MANIKARNIKA

Proactive Crowd-Sourcing for Location Based Services

By

Anuradha Vaidyanathan

A dissertation submitted in partial satisfaction of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY of Canterbury at Christchurch

Committee in charge:

Professor Harsha Sirisena **Chair**

Professor Mark Billinghurst

Associate Professor Wendi Heizelman

Professor David Johnston

Abstract

This thesis presents the design and evaluation of the *location* of a cell phone user, to enable more effective performance monitoring. One of the end-uses I propose is in emergency management, by means of a framework that distributes its functionality between establishing data-set characteristics that are relevant to the problem and a visual tool to evaluate resource-scheduling proposals.

Manikarnika is a modular framework, which finds translation in a prototype for Reverse 111. The first steps in the process were to establish whether the parameters I hypothesized as useful, indeed were. Using a statistically significant amount of traces, obtained from real calls placed on the network, the utility of the location metric was established. In order to investigate utilizing a second metric of reputation, a benchmark for evaluating ideas from Social Networks research was proposed, in order to move from arbitrary testing to a more systematic environment.

This dissertation details the measurement, design and evaluation of an end-to-end and modular framework for Emergency Management, where the functionality is distributed in order to easily incorporate the changing parameters of sources of information, emergency events, resource requirements of these events and identifying callers that might be able to provide better insight into a situation that is essentially very dynamic.

The chasm between research proposals for various end-uses and the application of the same to real life is one that I have tried to bridge in my work. By incorporating pieces from core Electrical Engineering measurements and simulation and extending the use of what was originally a tool built for training Emergency Responders to analyze various resource scheduling agents, which take into account a diversity of administrative domains, I lay the ground work for one possible solution, Reverse 111, which proposes the use of proactive crowd-sourcing for emergency response, with easy extensions to commercial location-based applications.

கற்றது கை மண்ணளவு கல்லாதது உலகளவு

ஔவையார்

Table of Contents

List of Figures

List of Tables

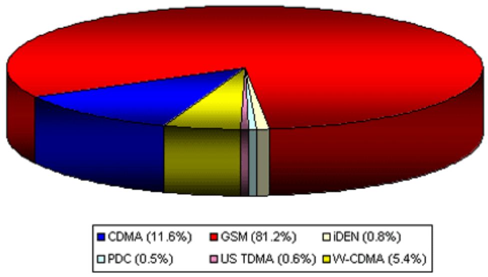
Acknowledgement

Its close to 10pm on a Monday night as I write this. I believe my work-ethic comes from watching two of the most hard-working people, my parents. They have always instilled in my brother and me, the quality of persistence, humility and resilience. The path to this degree has not been easy but, it has been very adventurous and it has taught me a lot about myself and the extents I am able to go with my body and my mind. I have trained to be an elite athlete in the 26 months I have taken to complete this degree and balancing several hours of training with research have been anything but trivial. I believe that desire cannot be manufactured, it has to come from within. Additionally, I have had the love and support of so many people that it makes me very happy to be naming a few of them in this section of my work. I would like to thank my mom and dad, Alamelu and Vaidyanathan, for being my blue sky. I would also like to thank my brother Siva, for his quiet ways and brilliant work-ethic. In many ways my baby brother inspires me to perfect my work and hone my skills. My friend Rowan Hooper has been instrumental in helping me keep my head together with the massive training and demands of graduate school. He is a legendary athlete and one day, I hope to be somewhere in the same zip-code when we are running together. His dad George Hooper has been a source of laughs and kind words as have my peers in the lab, Christina Dicke, Julian Looser and Mohammed Obaid. Several folks have helped me through the small and the big hassles including moving (Nate) and my love-hate relationship with my computer (Leigh Beattie). My advisor Mark Billinghurst has paid me to do my job and has been an inspirational employer and his working hours and constant advice are an inspiration. My other advisor Harsha Sirisena has entertained dozens of phone calls in the process of writing my thesis and has been very instrumental in the administrative overheads being seen to. Tom Furness has been an inspiration to read about and learn from and a very warm host when I visited his labs in Seattle, in 2009. I would like to thank Mr. Kapil Kapoor, a very special person who gave me a very big boost in 2009 and his colleague Mr. Kannan for helping me see the possibilities in my dreams of being an athlete. Thanks also goes out to Katy Bang, the super-efficient and ever-smiling person in the HIT Lab, who has made all of our lives so much more enjoyable with her energy. My staff in my Bangalore office, specially Kalavathi Krishna, has been amazingly supportive in this journey of mine, I still have a lot to learn from this girl. My recently re-discovered friend, Ashwin Mallipatna and his band of crazies have encouraged me to join in their efforts to start a non-profit organization for kids with Retinoblastoma. We are the sum of all our experiences so, there are a tonne of people that I have met along the way who have added to (said additions being normalized or otherwise) the process and the result, specifically Mike Dahlin and Scott Molina, who have been lighthouses in my life. Id like to take a moment to thank all of those that I have left out for I believe life is a zero-sum game and I hope to retain my good attitude and clarity of thought, while never feeling entitled to anything, except a fighting chance.

1 Introduction

The ubiquity of mobile phones has been enabled by three factors: (a) The establishment of standards world-wide to enable equipment manufacturers and network managers to provide cohesive systems that cater to different kinds of users (b) Process technology improvements that have made the shift from the “bricks” of the old days to the small and powerful cell phone instruments of the present and (c) User-Friendly Features and applications, starting with the simple Short Message Service to the more complicated Location-Based Services that render the cell phone as integral as a person’s house-keys or wallet.

The proliferation of mobile phones has grown in leaps and bounds, shifting from a total of 1.5billion GSM subscribers in 2005 [1] to a whopping 4 billion plus subscribers in 2009. The main contributor to this growth has been the adoption of this technology within the Asian countries, such as India and China. The total human population in the world is around 6.8 billion people, which might indicate that nearly 60% of the world carries a cell phone. These statistics are summarized in Fig. 1-1. More than 400 million mobile broadband subscriptions [2] indicate that around 10% of this population accesses the Internet from their cell phones. These trends have major implications for providing cost-effective commercial services and also providing a launching pad for new applications, such as security, in the case of emergencies.



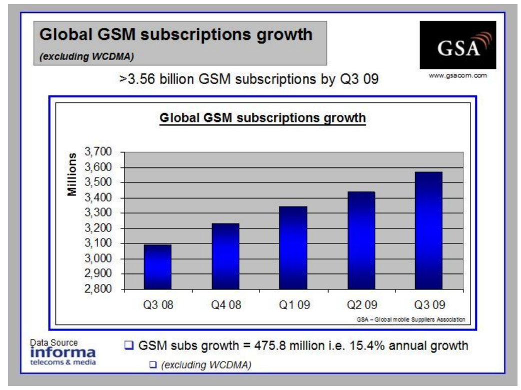


FIGURE 1-1. **Proliferation of Mobile Phone Technology.** This data was obtained from www.gsacom.com. The first figure indicates the most prevalent technology to be GSM and the second figure shows the growth in subscriptions in 2009.

A host of value-added services including mobile banking (one of the oldest services in this space), gaming, mobile health, multi-media (video and music) and accessing the web on the go have found an enormous audience and commercial vendors to drive their growth. Besides commercial and third party vendors, Telecom operators too have ventured into this space, thereby making it a very versatile and price-driven market. Beyond the provision of roaming facilities, which is directly linked to how a customer is billed for their usage, the concept of *location* is relevant in the context of mobile health and possibly newer applications such as Location-Based services and Social Networking, which are slowly gaining ground. The historical examples of the uses of location that made the headlines include one of some German climbers getting stranded in the French Alps, who were located after France Telecom traced them by triangulating on their GSM cellular radio-calls. Other examples include that of a lady sending an SMS from a boat that had lost power, from somewhere in Indonesia, to her boyfriend in England and getting rescued [3]. More recently, CNET reported a quorum of hackers who joined forces to make the lives of emergency workers more technology-savvy [4].

Although commercial and security applications take into consideration the use of cellular phones and mobile broadband as enablers, they have been slow to address the problems of scale, the continuum of an emergency and the benchmarks used to address the effectiveness of their proposals. The scale of an emergency varies with where the emergency occurs. For example, an emergency in New Zealand cannot be compared with an emergency in India, which is a country with a much bigger human population. With the base of mobile phone subscribers being most significant in Asian countries, any successful commercial or non-commercial application will need to address these issues, in order to cater to these subscribers. Current proposals limit themselves to certain types of handsets with pre-built features, certain assumptions about the network and are very heterogeneous in terms of applicability.

My thesis is that the use of *location* is central to enabling solutions that address the issues of scale, in providing meaningful solutions in both commercial and non-commercial applications, on the most ubiquitous *sensor* of them all, the humble phone. I propose the use of location to enable an emergency response service called *Reverse 111*, which extends beyond a random mass-phone call or mass-SMS, to utilizing the user's *location* to identify the users in the affected area, for more efficient information gathering and response. I further propose an overall *framework* for emergency response, Manikarnika, which takes into account the continuum of the emergency and addresses the relevance of the solution(s) to every step in that continuum. I hypothesize that this approach can enable three benefits compared to older proposals:

(A) **Utilizing the concept of location from ubiquitous sensors, cell phones** for emergency services. Considering the fact that a majority of the human population either already owns or will own a mobile phone in the years to come, I propose using these as *sensors* to locate people that are co-located with an emergency (or an event, for commercial applications and performance monitoring) to ameliorate and assist users. Emergency response systems today rely heavily on stratified networks [5,6,7] for responders and seldom take into account the importance and utility of a common citizen, who is on the ground that can provide valuable information to the responders. Furthermore, the solutions today are applicable at various stages of an emergency and don't use cell phones to the extent I propose in my system. The continuum of an emergency may be viewed as including several phases including the stage of occurrence, information and data-gathering via surveillance or more proactive or reactive methods, prevention of the emergency and last but not least, responding to the emergency.

User-mobility and location are critical to identifying citizens or subscribers that are co-located with the emergency, in order for the emergency responders to contact them, in order to get more information. By staying handset and technology agnostic, Reverse 111 can utilize spatial locality to gather information, prevent and respond to emergencies. Over and above simply locating users, understanding the scale of the emergency to determine the kind of information to gather from the ground, is mission-critical. An emergency in Christchurch is on a different scale from an emergency in Seattle, which is one, a different scale from an emergency in downtown Mumbai. The metrics to assess the effectiveness of proposals also changes, based on the scale of the emergency, which this system takes into account.

(B) Understanding how performance bottlenecks can be addressed from a network services-provider's perspective is the second aspect of my system. Provisioning networks for maintaining Quality of Service (QoS) guarantees are crucial from a management perspective. Given that users tend to move differently, at different times of the day, or during certain events within a locality, the provisioning within cellular towers has to be adjusted for peak performance. Load-Balancing and Cell-Breathing are two effects that I address in my proposal, in order to enable a performance engineer to reduce the number of dropped calls. By characterizing the data set effectively in a controlled population, such as New Zealand, I draw out the base features of live calls that can be used to simulate caller movement and caller traffic in studies that are simulating emergency situations. For example, if the average calling time and the average distance a user tends to travel while making a call if modeled accurately, the simulations of emergency situations have a more realistic flair to them, as the length and distance of calls and callers have been derived from real data.

Older proposals in this domain are heavily reliant on simulation-based studies that are at best hypothetical in their utility and result-set. By utilizing Per Call Measurement Data, which records live calls on the networks of New Zealand's leading mobile network and services provider, I present the effects of call-load during real events within major cities in New Zealand. In order to expose the effects of cell-breathing, I choose to focus on cell-sites that are identified to have consequential coverage and subscriber metrics, that make them good candidates to derive real-time effects of cell-breathing. This impacts provisioning directly as my system is able to predict overload and recommend addressing the situation as and when a performance engineer requires it.

(C) **Establishing benchmarks for commercial applications** is important in order to validate the proposal. In the case of a very important commercial application, Social Networking, I propose a benchmark, TESS, in order to establish good research practices. Social Network research relies on a variety of data-sources, depending on the problem-scenario and the questions, which the research is trying to answer or inform. By analysing some of the data sources indexed by the sizes of these data-sets and relating them back to the research question, which the data-set is used for, I assign a confidence metric to the data-set when tied to the field within the Social Network analysis that the data is used for. This will lay the foundation for benchmarking the results of any social networking study by means of concrete requirements for the data sets used in research.

To support my thesis, I make the following contributions in this dissertation:

1. I demonstrate the utility of the location metric, based on a ubiquitous sensor a cell phone to address a host of commercial and non-commercial applications, from network provisioning, to emergency response.

2. I evaluate a number of approaches to live emergency response, wherein the cell phone is an integral component, while addressing the scale of the problem that concludes that a greedy approach to emergency response outperforms other approaches. I propose the application of what can be considered crowd sourcing for the domain of emergency response, which is the first proposal of its kind, to my knowledge.

3. I describe a benchmark for a commercial application, Social Networking, which is one of the applications enabled on a mobile phone, in order to establish good research practices in this area of technology. I propose that social networks also extend to the emergency response aspect of my work, by establishing a notion of reputation.

4. I tie these elements together in a framework for emergency response, Manikarnika, and design a prototype for Reverse 111.

1.1 Design and Experimental Framework

I have developed a framework for commercial and non-commercial applications described here within by utilizing live call data termed Per Call Measurement Data

(PCMD), which is a feature of the Flexent switches, that are in use in the Alcatel-Lucent networks, within New Zealand [8]. These are the network managers for Telecom New Zealand, which has one of the largest subscriber bases in the country. In addition, I have used data from other sources, including PSMM, PERISCOPE, etc. These data sets are described in more detail below. The PCMD feature with Alcatel-Lucent, collects selected call measurement data for every call. The collected data includes call technology details, resource usage, and call failure information. The PCMD feature collects data for complete and incomplete calls (for example, denied calls and setup failures) so that data for these call attempts can be analyzed by other tools or post-processing systems. The Vallent Prospect System consists of performance analysis tools for monitoring, fine-tuning or troubleshooting an existing Flexent/AUTOPLEX Wireless Network. It includes modules to enhance the operation, maintenance, and engineering of the wireless system. Modules can be used to monitor system performance, to report system configuration, and to diagnose, track, and solve many types of wireless engineering and system operation problems. With a single software product, Vallent Prospect supports all access technologies (AMPS, TDMA, and CDMA) on both cellular and PCS systems. RSSI data used in this design was collected from the Prospect tool. PSMM data is part of the PCMD data-set but can only be collected if the feature is turned on, which I did for my studies.

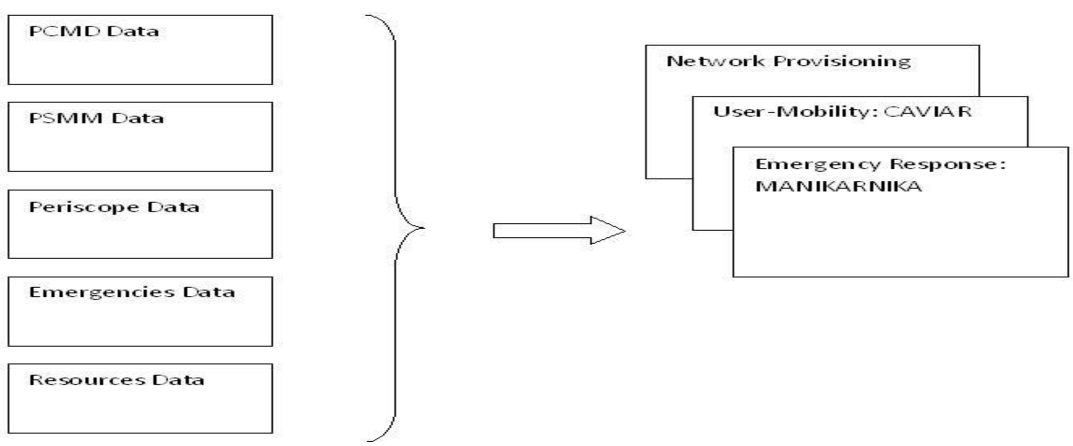


FIGURE 1-2. **Design environment.** Five sources of data are used in my environment, PCMD, PSMM, Periscope, Emergencies data and resource location data. These form the inputs to my end-applications for Network Provisioning, User-Mobility tracking and Emergency response.

Periscope is a tool that integrates information about every caller’s handset and subscriber details, which I propose using for evaluating some ideas on reputation. Emergencies data have been collected from public information web sites, specifically regarding statistics on locations and frequencies of emergency events in the greater Seattle area. Resources data includes data about the physical location of fire stations and police departments to give my simulation tool a flavor of reality.

Maximum likelihood methods are used in the geo-location algorithm in order to refine triangulation estimates to the highest probability location. Field calibration has a per-call median error of ~140 meters, averaged over all call locations wherein accuracy is increased when the calls are placed closer to cell sites. For this study, data from two distinct switches were collected and a sample of that was anonymised and used as input for characterizing various aspects of the network. I also used the Haversine formula [9] to calculate the distance of the user's movement in the duration of the call.

1.2 Overview of the Dissertation

This thesis is comprised of six main chapters. The first provides the groundwork and sets the stage for my design choices and Emergency Response proposal. The remaining chapters provide more details about my choice of *location* as a **valid** metric for evaluating caller traffic and obtaining information pertaining to the changing nature of an emergency, provide some reasoning into providing accurate benchmarks for Social Networks (which I could use as a second metric to establish caller *reputation*), propose the use or scheduling agents to evaluate different responses to emergencies with the goal of optimizing response actions across varying scales of emergencies and administrative requirements and tie these back into a simulation environment, which enables evaluating some of the ideas with visual aids.

In Chapter 2, I review key technology trends that affect my design choices when proposing the use of *caller location* based on a user's cell phone, as a reliable sensor to gather information about emergencies. By providing perspective into how ubiquitous a cell-phone is, across four generations of mobile telephony, I explore the use of being able to locate a caller that is somewhere close to an emergency by taking examples from commercially successful ventures into Location Based Services and more *en masse* proposals for Emergency Management, including Reverse 911 and E-911.

In Chapter 3, I introduce the use of the *location* metric for a variety of performance measurements relating to traffic management, load balancing and cell breathing. These studies are used to motivate the use of a caller's location to obtain information about an event that they are co-located with. In my analysis, certain characteristics are established regarding the nature of calls, the average distance users travel while making calls, the average length of their calls and the variety demonstrated in a trace, which that is selected for a study. These parameters are important to establish in order to set the bounds for my experimental and trace-driven framework. New Zealand has a relatively small number of cell-phone subscribers (as compared to Seattle or India). This is a controlled data-set and provides key insight into how the data-traces must be analyzed and segregated, in order to focus the results on the parts of the data that matter, rather than the sum of the whole. In addition to validating older work on known effects such as cell-breathing *with live-call data* this chapter leads to my validation of new hypotheses around load-balancing across cells in a mobile-phone network. By using a hybrid of PCMD data and Prospect data, I analyze the effects of caller-traffic moving in a certain direction, when a Rugby game is being played, in order to establish an observation about how call-traffic really gets absorbed. In case of emergencies, human beings tend to move in clusters *away* from the emergency, along established exit-routes. Therefore, understanding how the underlying network behaves is crucial in order to select the right set of callers to establish contact with.

In Chapter 4, I present the deficiencies with Social Network research data sets. Ideas from Social Network research could be applied in my case, in order to establish a notion of *reputation* amongst callers. The establishment of reputation is not central to this chapter and the focus of the work laid out here is in commenting on the desirable properties of data sets when performing research on social networks. Data sets that are too small, have not been sampled over a reasonable amount of time, do not reflect the expertise of multiple types of participants and have been obtained from unreliable sources do not provide much *confidence* to the ideas that can be adapted from this field of work. In this chapter, I propose a benchmark, TESS, to evaluate the strength of the ideas in this domain with more robustness.

In Chapter 5, I present *scheduler* that can be used to address how different users, administrative domains and geographical domains can interact to effectively schedule resources across emergency situations. Three different schedulers, a greedy scheduler, a lottery scheduler and a round-robin scheduler are proposed to evaluate different scenarios that might lead to different results when dealing with emergencies. I evaluate these schedulers based on throughput and latency that they demonstrate while addressing the emergency situations created for them.

In Chapter 6, I present a modular framework *Manikarnika* that takes a step-wise approach to evaluating how emergency management can be addressed. In the case of emergency situations or disasters, there often exists a *continuum*, which refers to all the events before the emergency occurred, during the emergency and the response actions to mitigate this, after the emergency has occurred. In order to effectively manage emergencies, it is imperative to understand which portion of the emergency continuum my solution is addressing. Further in this section, I also demonstrate the use of AR to overcome linguistic barriers within emergency response, in order to demonstrate that cell-phones have the technological ability to apply proposals across different cultural and geographic barriers, if developed correctly. The recent interim reports on the appropriate warning systems for bushfires in Victoria and tsunami alerts in New Zealand, motivate the utility of the cell phone for emergency management and response.

In Chapter 7, integrates all the pieces of the design addressed in chapters 3-6 into my prototype for Reverse 111. I show the user-friendly aspects of my tool, which can be dynamically configured to take into account various aspects of an emergency including how many events occur, the frequency of occurrence, how many resources are available, how many administrative domains are involved in that run of the simulation, how many pieces of information are made available for the person running the simulation to view, etc.

Lastly, Chapter 8 summarizes the key conclusions of this dissertation, points out some of the lessons I have learned in my various exercises and discusses future work.

2 Technology Trends

This chapter describes several key technology trends that will influence the design of mobile phone applications in the coming years and motivate the relevance of these trends to Emergency Response. Section 2.1 outlines the basic proliferation of different kinds of cellular phone standards, handsets and hardware parameters. These trends affect the compatibility of various applications across various demographics. Section 2.2 describes the various applications specifically for the mobile phone market that are gaining ground. These include the traditional mass markets of mobile banking, gaming, etc. and also the emerging applications of location-based services and emergency response. Section 2.3 explains how these technology trends impact specific aspects of my framework to provide proactive rather than reactive solutions to emergencies. Section 2.4 addresses other technology trends that are not central to my application, for the sake of completeness.

2.1 Technology Trends

In year 2002, a set of papers detailing records of the original GSM task force were compiled. In this set, a study by Josef Huber [3], methodically deconstructed the market model, traffic trends and the technology parameters classified by services, and offered predictions of the same between years 2005-2010. This study was carried out primarily to arrive at an appropriate methodology for spectrum allocation. These predictions are shown in Table 2-1.

Service Categories	**User Net bit-rate**	**Penetration urban year 2005 (%)**	**Rate environment year 2010 (%)**	**Applications**
High Interactive Multimedia	128kbps	0.5	5	Videophone, real-time services
High multimedia	≤ 2Mbps	5	18	WWW type services
Medium	< 384kbps	8	18	WWW documents, video

multimedia				streaming, internet/intranet access
Switched data	14.4kbps	10	10	e-mail access, data
Simple messaging	14.4kpbs	25	40	SMS, enhanced messaging
Speech	16kbps	60	75	Voice

TABLE 2-1. **Technology predictions for 2005/year 2010.** This data was obtained from Josef Huber's expose on Spectrum Aspects in the UMTS Related Work of the European Commission, UMTS Task Force, UMTS Forum and GSM Association.

Rural, urban, indoor and outdoor environments had to be treated as a whole with due consideration being given to population densities, as shown in Fig. 2-1.

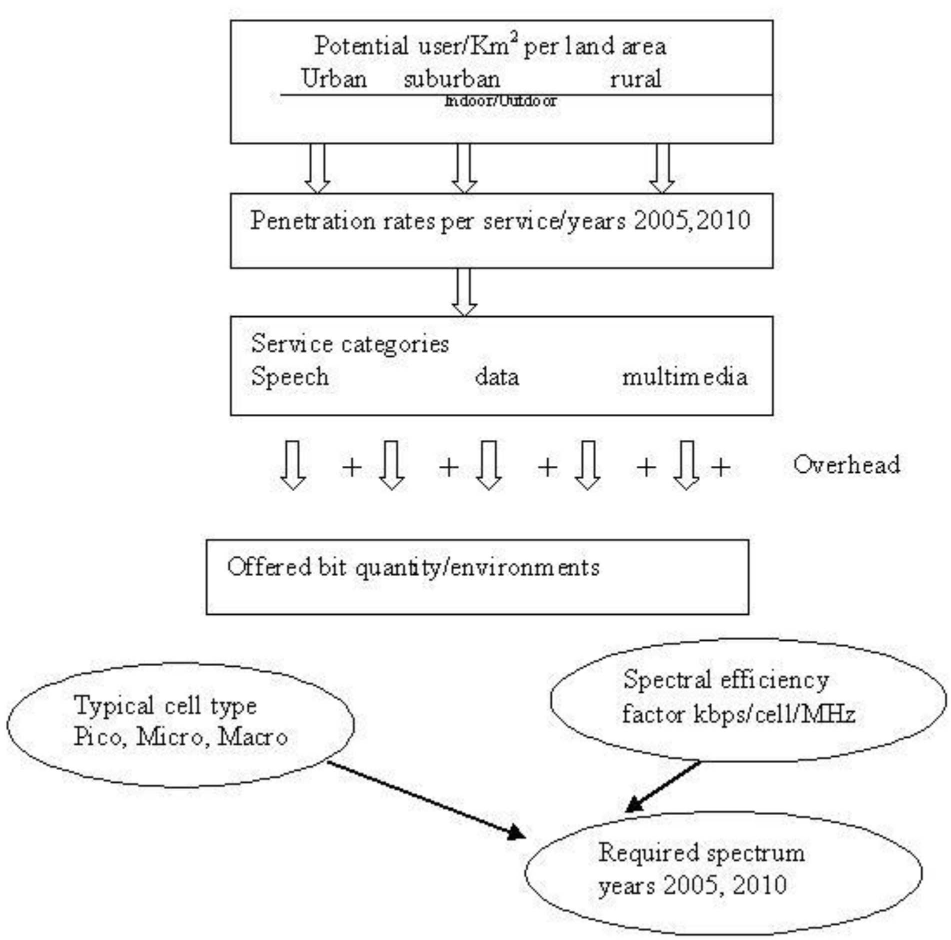


FIGURE 2-1. **Considerations for spectrum provisioning.** This data was obtained from Josef Huber's expose on Spectrum Aspects in the UMTS Related Work of the European Commission, UMTS Task Force, UMTS Forum and GSM Association.

Both usage statistics, as shown in Fig. 2-2 and technology parameters, especially bandwidth (owing to the increased role of packet switched networks) has far exceeded initial expectations and predictions. Furthermore, the success of such technologies as VoIP has in effect validated the push towards 4G networks, with a predominantly IP backbone.

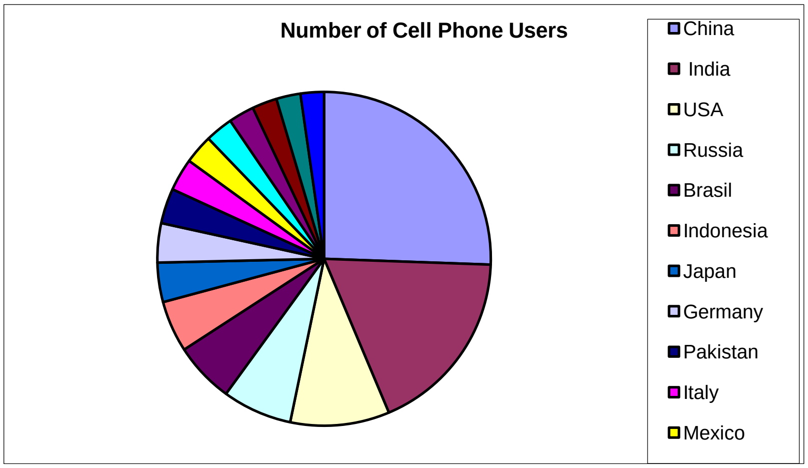


FIGURE 2-2 **Top 16 countries using cell phones.** This chart depicts the usage, in total number of phones, based on the top-16 countries that are using them. These countries were chosen to best represent the most populated parts of the world and covers most of the world except Australia/New Zealand, which don't figure in the top 25 even.

In addition to this, the least common denominator when it comes to features, even in low-cost handsets is pretty exhaustive, thereby enabling various new applications. I explain these trends in detail in the following sub-sections.

2.1.1 Four generations of mobile telephony

I present a historical perspective on the evolution of standards for mobile telephony over the years. This section is important in order to understand the *percentage* of users I am able to utilize in order to obtain information from these users at the time of an emergency. The multitude of standards poses a serious issue in proposing a solution for emergency management using cell-phones because the proposal has to address the diversity in standards. For a long time, in the Asian markets, even with the advent of 3G standards, subscribers were buying 2G handsets for reasons of costs [10]. This is no longer the case as the cheapest phone now comes with such features as a camera with good resolution and sufficient memory and thus increases the base of subscribers I can contact, in order to obtain information about the changing resource requirements to address emergencies.

The earliest version of mobile phones was the two-way radios, primarily used to communicate on set frequencies, not utilizing the telephone network. Between 1910 and 1973, the technology saw several steps in its evolution. In 1973, Motorola demonstrated the DynaTAC prototype, which was probably the first commercial version of what is today smaller and lighter by several orders of magnitude. These were still analog devices and the introduction of the 2G standard, which was the term used for TDMA and CDMA systems came about in the 1990s, to introduce digital, circuit switched transmission. The phones in this generation still weighed about 200 gms at their lightest and saw the introduction of several commercial services such as the short message service (SMS), downloadable ring tones and roaming. 3G networks saw the introduction of competing standards including WCDMA and CDMA2000. Although these standards had to stay cohesive with the IMT-2000 specification (standards of data-rates around 384 kbits/s outside and 2Mbit/s inside), they introduced a lot of heterogeneity. Intermediate standards such as 2.5G networks were also introduced to buffer the transition between 2G and 3G such that at the end of 2007, the subscriber base for GSM had reached nearly 300 million. Although the proliferation of mobile phone technology has far exceeded expectations, the data-rates on circuit-switched digital networks have not scaled as well. While the main driver for 2G networks was voice, newer applications such as video and digital television are predicted to be the drivers of what has been termed 4G networks, as the limits of short-range communication evolve. Table 2-2 summarizes the generations of telephony according to various parameters including services supported, standards, bandwidth and the core network.

Technology	**IG**	**2G**	**2.5G**	**3G**	**4G**
Services	Analog voice, synchronous data to 9.6 kbps	Digital voice short messages	Higher capacity, packetized data	Higher capacity, broadband data up to 2 Mbps	Higher capacity, completely IP-oriented, multimedia, data to hundreds of megabits
Standards	AMPS, TACS, NMT etc.	TDMA, CDMA, GSM,PDC	GPRS,EDGE, 1xRTT	WCDMA, CDMA2000	Single standard
Data Bandwidth	1.9 kbps	14.4 kbps	384 kbps	2 Mbps	200 Mbps
Core Network	PSTN	PSTN	PSTN, packet network	Packet network	IP

TABLE 2-2. **Four generations of mobile telephony.**

2.1.2 The proliferation of handsets

While network technology progressed in supporting higher data rates and more features, one of the greatest leaps in subscriber base occurred when the Asian continents started to engage with mobile phones. Given that the population density in this part of the world is unparalleled, two driving factors determined the success of both handsets and network technology, first movers and price. For a while, even though 3G technologies found an audience. Even though GSM was adopted early in India and China, the handset prices drove the subscriber base, in the initial years. There were many reviews that brought out the fact that even though 2.5G and 3G services were available to subscribers, the price of the handset forced them to choose 2G phones, thereby rendering a lot of the services redundant. Pricing services based on customer requirements has bridged this gap and instruments are more capable, even at the lowest prices, in the current market. Fig. 2-3 shows the handset market share, with the biggest players being Nokia, Samsung, LG and Motorola, followed closely by Ericsson [11].

Brand	**Units Sold (in millions)**	**Market share 1st Quarter, 2009**	**Market share 4th quarter, 2008**	**Market share 1st quarter, 2008**
Nokia	93	40.43%	41.93%	41.1%
Samsung	45.8	19.91%	14.54%	16.48%
LG	22.6	9.83%	7.44%	8.68%
Motorola	14.7	6.39%	6.51%	9.75%
Sony Ericsson	14.5	6.3%	7.6%	7.94%
Blackberry	7.8	3.39%	0%	0%
Apple	3.8	1.65%	0%	0%
Others	27.8	12.09%	15.32%	17.1%

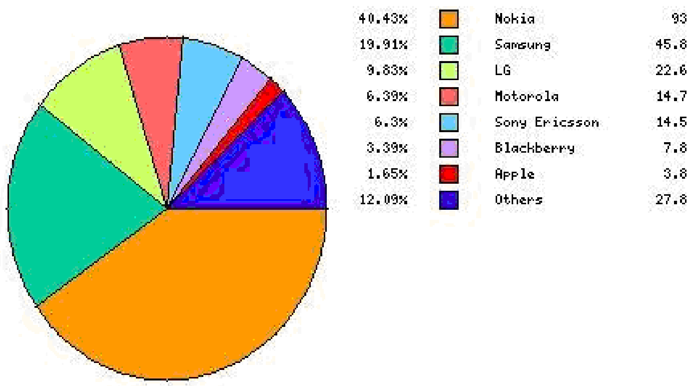


TABLE 2-3. **Handset Market share.**

Features of the handsets I am interested in include price, memory, image-resolution and ability to connect to the Internet. I profile these parameters for 42 handsets from the leading brands, based on various price-points [12,13]. I find that the most basic handset, still has a camera and is Internet ready. The price-points and concerns about majority of subscribers in the Asian countries using 2G instruments by reason of price are slowly being bridged. The extended results are presented in Appendix I, Table 2-4.

Handset Make	**Price**	**Memory**	**Image Resolution**	**Internet Enabled**
Motorola AURA	Rs.111,492	Memory Card Type - MicroSD; Extensible Memory-2 GB	240 x 320	WAP 2.0/xHTML
NOKIA - 8800 Sapphire	No result found	Internal Memory-1 GB	240 x 320 pixels	WAP 2.0/xHTML
SAMSUNG - Omnia 16GB	List price- Rs 41,572; Rs 27,895	Internal Memory-16 GB;Memory Card Type- micro SD;Maximum Extensible Memory-8 GB	400 x 240 pixels	WAP 2.0
SAMSUNG - Omnia 8GB	Rs.26,500	Internal Memory-8 GB;Memory Card Type- micro SD;Maximum Extensible Memory-8 GB	2592x1944 pixels	WAP 2.0/xHTML, HTML, RSS feeds
SONYERICSSON - W995	Rs.24,990	Internal Memory-118 MB;Memory Card Type- Memory Stick Micro (M2);Extensible Memory-8 GB	240 x 320 pixels	WAP 2.0/HTML (NetFront), RSS reader

TABLE 2-4. **Trends in the features of handsets .**

2.2 Trends in the Application Space

In this section, I discuss four major mobile applications and value-added services that have emerged and show promise for sustained growth including short-message services, mobile banking, gaming, health, multi-media web-browsing. Historically, the roaming, and SMS (as applied to banking) were the initial commercially consequent offerings on cell phones [3]. With the introduction of more powerful devices with more transistors, translating to increased computing power, more applications have evolved. Besides the main applications indicated above, with the proliferation of HSPA devices [14] , Location-Based services that enable mobile marketing and social networking sorts of applications are also fast emerging. Enterprise applications are a part of this development including automation and logistics. Location-Based services in particular require the introduction of new elements on the network (for resolution accuracy, orientation requirements and privacy requirements).

2.2.1 Traditional Mobile Applications

The earliest service for mobile phones was the short message service and the first industrial sector to take advantage of this was the banking sector, to enable customers to receive alerts about new rates of interest etc. The major mobile phone applications today include mobile banking, gaming, health, multi-media and web-browsing, probably in that order. Mobile banking also extends to brokerage, accounting and financial information services and are primarily based on transactions. Several studies by financial consultancies indicate that online banking is a trend that is one the rise and will see an adoption of nearly 35% in households, by year 2010. Upto 70% of call volume to banks are predicted to be from mobile phones. Contact less payments are the other trend that are expected to gain ground in the coming years. The incentive for banks to retain their customer base by providing a host of value-added services remains the primary driver for the heavy investments in the m-Commerce space. Fig. 2-3 shows the growth of m-Commerce within one specific market, Japan. The rates of growth are comparable world-wide.

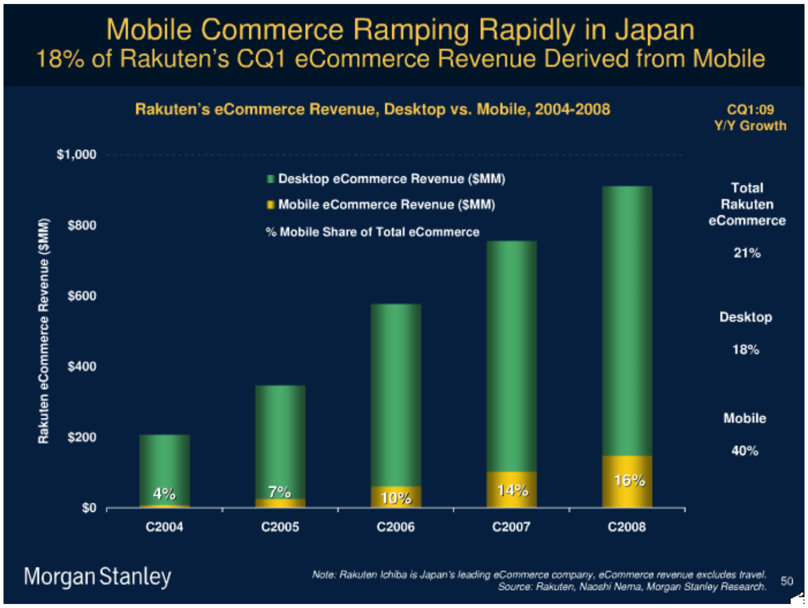


FIGURE 2-3 **growth of m-Commerce.** This chart depicts spread of m-Commerce in Japan.

Mobile gaming is another area that has also been around traditionally and is gaining ground with the improved processing capacities of handsets. I discuss mobile gaming in India, in order to motivate the importance of this market from the standpoint of one of the most dense subscriber bases in the world.

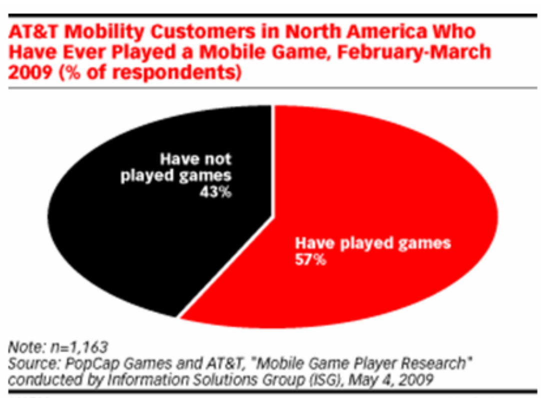


FIGURE 2-4 **Proliferation of mobile Gaming.** This chart describes the percentage of the mobile game players in North American markets.

Even though mobile gaming only enjoys around 1% of the value-added services market (as compared to SMS), the onset of 3G has lead to predictions of this industry crossing a couple of hundreds of millions of dollars, with an estimated growth rate of up to 200% in the 2010 financial year [15]. Based on the numbers collected in the first quarter of 2009, it was found that nearly 30 million urban Indians use their mobile phones to play games daily, which constitutes about 12.5% of the total subscriber base. The total number of players exceeded 100 million, with the frequency of use varying between once a week to up to three times a month as shown in Table 2-3 and Fig. 2-4

(Two month ended March 2009, Urban Indian Mobile Phone Users)

Frequency of playing games	Unique Users (Millions)
Played 1 to 3 times in last month	46.0220
At least once a week	28.5380
Almost daily	30.4700

TABLE 2-3 **Feb/March 2009 statistics on mobile gaming.** Nokia continues its dominance in Indian mobile space accounting for 65% of power user's handset followed by Sony Ericsson a distant second (10.6%).

Health-care services providers have started to rely on mobile devices to assist in various stages of the multiple procedures a patient and an administrator goes through. A recent study at Mr. Sinai hospital in New York [16] discovered a higher rate of compliance with medication doses amongst patients, when they were reminded regularly by text messages. Devices for physicians to record patient information and access the same locally and remotely have been well accepted in the North American and European markets. This area of mobile applications was probably the first to consider patient confidentiality as being a concern, when moving to such platforms. Encryption is thought to solve a few of the application-layer concerns but the issues of privacy are an active topic of debate.

Mobile multi-media traditionally revolve around music, as it is slightly easier to render this service as compared to streaming video. In the case of music, the progression has been from delivering a user's favourite songs as their ring tone using true tones, chaku-uta and chaku-uta full standards, typically recoded in the MP3, AAC or WMA formats. In 2004, ring tones represented what was nearly a $4 billion market, excluding the United States. Other forms of multimedia on phones includes video and images, which are more reliant on 3GPP technologies.

The mobile web is probably the biggest technology driver that has changed everything, from the way sites are built[17], the browsers used to access the web from a phone (Opera mini, google android, skyfire, etc.), the standards to support development efforts (shown in Fig. 2-5), etc.

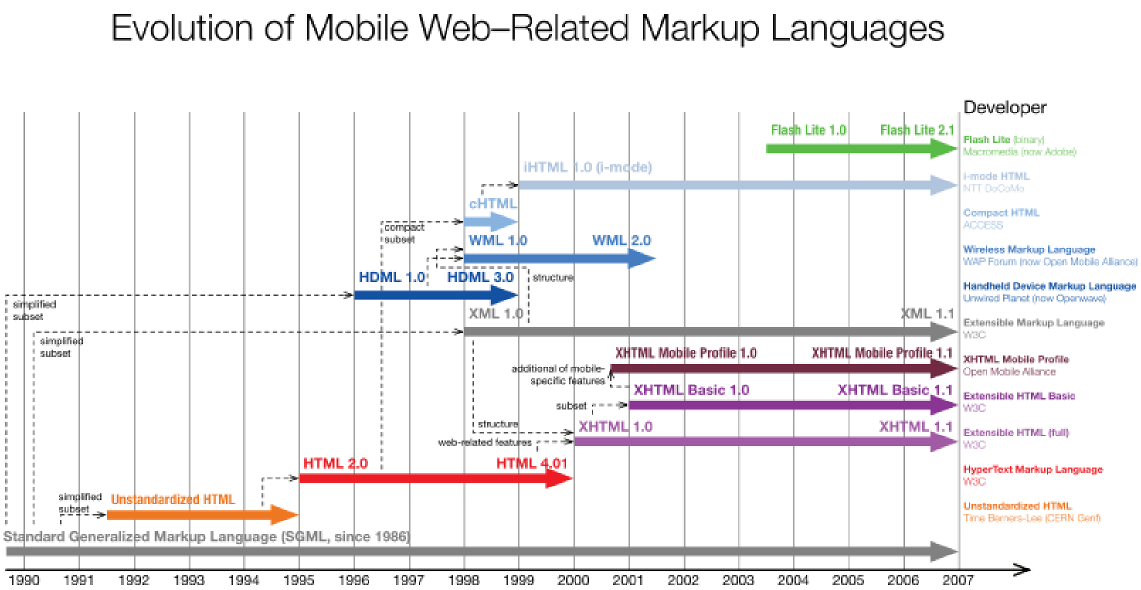


FIGURE 2-5 **Evolution in Markup Languages.** The enablers for mobile applications are increasing by orders of magnitude over the years.

2.2.2 Location-Based Services

Location Based Services trace their first instances back to 2001, when DoCoMo [18], a Japanese phone operator introduced triangulation to locate pre-GSM handsets. In order to personalize services, LBS is a new set of applications that take into account the exact location of the user, in order to suggest a host of useful co-located facilities such as the nearest ATM, their friends who live in the area, the closest place to eat, etc. Commercial LBS implementations are slowly gaining ground with handset manufacturers, mobile phone operators and third party application developers with the introduction of GPS receivers in the chipset, programming environments such as BREW [19] and the commercial success of such applications as social networking. A number of methods can be used to locate the users including control plane locating, wherein radio signal delay of the closest cell phone tower is used to find the phone, GSM localization, wherein the relative position to the cell-site is once again used to find the device and finally, near LBS technologies such as Bluetooth, WLAN, infrared, and RFID are also options. In order to address user-privacy, several legislative acts have been put into place, requiring the user to *opt-in* by means of an SMS or some interface with the carrier or service providers.

2.2.3 E-911

Enhanced 911 or E-911 [20] has been introduced in North America wherein the telephone operators are required, by law, to enable enhanced 911 services on both wireless and wire line phones. In the case of mobile phones, the operators have been directed to only cater to location capable devices, such that when a user places a 911 call from any of these devices, the callers phone number, cell phone tower, latitude and longitude are relayed to the Public Safety Answering Point (PSAP), upon request. The allowed resolution, per FCC regulations is between 50-300 meters. Various methods may be used in order to realize these requirements and the determination of location of the handset is usually achieved by recording one or more of the **Angle of Arrival,** the **Time Difference Of Arrival (TDOA)** or what are referred to as **Location Signatures.** The angle of arrival approach requires at least two different cell towers, wherein a caller is located depending on where the lines along the angles of the tower intersect. The TDOA approach utilizes multilateration wherein the networks use the time difference across towers to determine distance. These two approaches are based on a line of sight (or signal acquisition) whereas the approach based on location signatures stores recall patterns such as multi-path, which mobile phone signals are known to demonstrate depending on their location.

2.3 Implications for system design

2.3.1 Proactive Crowd Sourcing

Most proposals for managing emergencies, except in the areas of surveillance, are *reactive* in nature. This curbs their effectiveness by several orders of magnitude. *Crowd Sourcing* is a term used to describe de-centralizing operations and trusting the work to the co-ordinated actions of a crowd. In the case of emergency management, a need of the hour is to enable more *proactive* solutions. Instead of an Emergency Services provider waiting until the emergency has occurred, assessing the nature of the problem and then deploying resources, if they could instead rely on citizens co-located with the emergency, the management of the emergency would become more agile and effective. Users co-located with the emergency are, arguably the best sources of information on what might be needed to mitigate the effects. Sometimes, the requirements are not intuitive and no amount of surveillance or pre and post-op training can predict the needs of the people affected by the disaster. For example, in the case of Hurricane Katrina, one of the biggest requirements that the responders fell short of was ice! [21] Both E-911 and Reverse 911 [28] fall in the category of reactive systems or approaches to the problem. In the case of Reverse 911, the biggest barrier to effectiveness is that in case of a fire in San Diego, locating all households in San Diego *en masse* may not be the most efficient plan of action. There might potentially be several people that have travelled to San Diego from surrounding areas or from afar and locating them is best achieved via a cell phone, rather than fixed geographical and list-based approaches.

The notion of crowd sourcing can also extend to other aspects of emergency management. Consider a situation where two adjoining counties (or administrative domains) are involved in an emergency. In order to best handle the situation, it might be beneficial for the two counties to work in tandem to share their pool of resources and allocate resources based on the shortest path the site rather than dealing with bureaucratic limitations. These ideas have been evaluated in my work through a number of schedulers used by first responders.

2.3.2 Location-Based Management

Although location-based services have been discussed extensively in the commercial arena, there are sufficient enablers that make a strong case for extending this metric of location for emergency management. As described in section 2.3.1, the proactive nature of my framework, Manikarnika, discloses a method to contact local citizens, on the ground, that are co-located with the emergency or a disaster, to get their input and feedback on the situation. Enough technology, standards and privacy discussions have been carried out, from the contributing actors in other areas of mobile phone applications such as banking, gaming, health-care and E-911 that translating these to provide *proactive* location-based management of emergencies is not too much of a leap. A second notion of *reputation* could be introduced, in order to increase the level of trust in a citizen's reporting and I explore associated benchmarks to enable this idea in the coming chapters.

Communication for emergency providers has long suffered from *interoperability issues* and several mandates address the same [5,6,7]. Some of the KPIs worthy of note include resiliency, accessibility, portability, interoperability, expandability and affordability. The sheer number of devices, protocols, network requirements and other issues call for a simple parameter, namely *location* to address the vast sea of incompatible modules.

2.3.3 End-to-End design

The scale and severity of emergencies vary with the location in which they occur, the nature of the emergency itself and the population density, in case the emergency affects a set of people. Several examples come to mind from news reports including the Tsunami in the Indian Ocean, the Mumbai terrorist attacks, Hurricane Katrina and the Air New Zealand Flight 901 crash on Mt. Erebus. These examples typify different locations, population densities and communication patterns and requirements in the span of their occurrence. In most cases, the impact of a disaster or emergency is measured by the number of lives affected by it.

Cell phones can be used to pin-point the exact location of users across a region and in the case where this mode of communication is still viable, this is a very important piece of information to mitigate the emergency. In the case of large-scale natural disasters, I might encounter the case where the cell-phone towers themselves are no longer available, rendering a sizeable coverage hole in the network. In this case, the most proximate communication region is important to establish to both gather information from users that are able to communicate back with the emergency management team about what they see in terms of the disaster and to possibly evacuate other population groups in the vicinity. In cases such as the co-coordinated terrorist attacks in Mumbai and New York, the cell phones become very relevant as reliable sensors. In both cases, a series of attacks were co-

coordinated and carried out in tandem. The scale in these two situations is different as Mumbai is far more populated than New York. However, the basic utility of the cell phone remains indisputable. Furthermore, the metrics that determine the success of a solution or proposal will vary based on where the emergency occurs. For example, if there is an earthquake, in the case of New Zealand, the metric to be addressed would be how quickly the resources can span a certain distance to get to the affected area vs. in India, the metric would now be the number of lives (given the higher population density) that the dispatched resources can save.

I propose addressing the continuum of an emergency pinpointing the exact stage where the solutions are being applied. This enables the first-responders that are not yet out on the field, in obtaining information from civilians (whose location has been determined using caller-records) that are co-located with the emergency and adds to the resiliency requirement of the communication infrastructure.

2.4 Other related trends

2.4.1 Augmented Reality for Advertisements

In the area of mobile marketing, the use of augmented reality has gained ground and been commercially viable [22]. I present the utility of Augmented Reality in Emergency Response situations, in order to provide efficient and expeditious information, while addressing language barriers that can be used to manage the emergency. Emergencies vary in scale and severity and depending on the actual location of the emergency, both the impact of the emergency and the response that is enabled, affect the quality of managing the situation. Linguistic and other communication barriers present themselves on a magnified scale in times of an emergency. The simplest aids to civilians assisting emergency managers to relay information from the ground would be visual. I briefly explore this alternative by means of proposing an extension of Reverse 111 that utilizes AR to provide visual cues to civilians and establish a two-way communication between the emergency response center and civilians co-located with the emergency, thereby presenting a case for AR in Emergency Response. Given their popularity in Advertising, it is not a far leap of faith to use the same enablers for Emergency Response.

2.4.2 The use of multi-player games for training

The ubiquity of using games and simulations in various departments of education necessitates a deeper understanding of the various components of these simulations, their similarities and differences, in order to (a) effectively apply the lessons learned in this field to other fields and (b) understand the necessary components or abstractions within simulators used for education. Several taxonomies have been proposed in order to demarcate the functional and design-oriented aspects of prolific games, that are in use today. These games may be used for a variety of purposes, other than the traditional goal of entertainment. Several studies have been conducted in order to propose semi-formal and formal [23, 25, 26] methods for game-design. Game classification and game design have also been conducted [24,27] in an attempt to categorize games based on concepts and representations of space, time and other variables within games. My framework for Emergency Management and associated prototype is built as an online, multi-player game, wherein one or more administrative domains can partake in the exercise of emergency resolution. The schedulers that are evaluated in my work are played in the system in *oracle* mode where only one player, the system, is evaluating the various aspects of scheduling resources for emergency events.

3 Location

Understanding how users move while using their mobile phone allows network engineers to perform congestion control, proactive customer servicing, efficient resource partitioning and error handling at the various call sites. Telecommunications services and associated networks generate vast amounts of data on a day-to-day basis. It is important that these characteristics be monitored in order to provide various QoS guarantees and maintain a loyal customer base, by providing on-time and proactive services. As the number of subscribers increase, the importance of *proactive monitoring* and *error correction* become more necessary. In the case of handling emergencies, understanding how populations move either during the course of a day or aggregately during special events, will help elucidate the power in utilizing co-located citizens to provide information about emergencies. Previous studies have used models of cell phone data in order to simulate scenarios to evaluate various metrics and propose improved solutions to aid performance. Accurate models need to be validated by real-time data and these models further need to characterize their data sets along the patterns seen in real-time data. Using **P**er **C**all **M**easurement **D**ata (PCMD) from a leading service provider in New Zealand I analyze various factors including the mobility exhibited by the subscribers, load-balancing on the network and cell-breathing. Section 3.1 outlines the methodology used to perform my experiments. Section 3.2 presents my findings on user-mobility, in order to motivate the importance of the *location* or *position* metric. This section talks about the basic findings relating to how far a user is likely to move in the duration of a call, the number of cell-site changes they are likely to make, etc. in order to describe the basic characteristics of callers and the data-set itself. Section 3.3 narrates my findings on Load-Balancing, based on callers moving on the event of an All-Blacks Rugby game in Wellington. This data is related back to caller traffic and patterns of movement on regular weekdays and weekends to motivate how understanding a user's movement and location can make for better provisioning of networks. Section 3.4 discusses my findings on the phenomenon of Cell Breathing wherein calls made at the edge of a cell are often lost as the load on the cell increases. Section 3.5 outlines the related work and 3.6 presents my conclusions.

3.1 Methodology

In this section, I describe the steps of collecting data, processing it for detailed analysis and characterizing the data to pick the elements that are relevant to the study being performed.

3.1.1 Data Collection

For purposes of this study, traces were collected from CDMA networks, where a Per Call Measurement Data (PCMD) feature is present. PCMD [8] provides access to key network information for every 3G1x (voice, SMS or data) call that is placed via the network. The data recorded pertains to several aspects including Identity (MIN, ESN), service type, number dialled, call length, signal quality, timing from pilots, sector in which the call was placed, the latitude and longitude of the cell tower where the call commenced and ended, call result and cause of failure etc. PCMD records provide an unprecedented, unobtrusive view of customer behaviour and aspects of network performance. In order to obtain a geographical view of information pertaining to the network, a geo-location algorithm has been used to extract accurate location information from PCMD. The timing and hand-off systems, unique to CDMA, alongside accurate information about the network allow the user location data to be gathered easily. Maximum likelihood methods are used in the geo-location algorithm in order to refine triangulation estimates to the highest probability location. Field calibration has a per-call median error of ~140 meters, averaged over all call locations wherein accuracy is increased when the calls are placed closer to cell sites. For this study, data from two distinct switches were collected and a sample of that was anonymised and used as input for characterizing various aspects of the network.

PCMD traces are collected from three switches in Auckland (AK), Christchurch (CH) and Wellington (WN), in order to analyse user patterns in these big urban centres. The data is accumulated at a file server and transferred there via FTP. This is then sent to a data-loader with my custom software, which parses the raw PCMD to scan for characteristics of interest. The data-loader contains means to load the data into custom databases, run queries on the data and means to interact with the data and feed it to a visualization interface. Standard maps of New Zealand in general and three urban centres in particular are also stored in the data-loader, in order to map the location of calls and track them through the lifetime of the call. Fig. 3-1 shows how the data is collected from three different switches and sent to an FTP server, where it is temporarily stored before being passed onto a data loader and processed for later use.

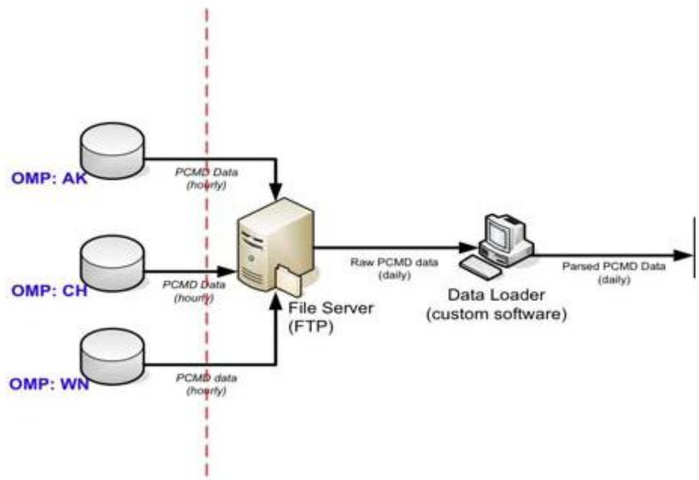


FIGURE 3-1. **Data Collection from three switches.** Data being collected by the telecom services provider at three locations, Auckland, Christchurch and Wellington.

Table 3-1 shows some of the fields in the PCMD data set that demonstrate the utility and relevance of this data to my studies.

Field num.	**Length**	**Verbose output field name**	**Field description**
1.	16 bits	PCMD Version	Identifies data layout. Unique for each release.
2.	32 bits	Sequence Number	Sequence number that is incremented for each call record

3.	32 bits	Call start Time	Time call was started Set according to the access or page response message as follows: 1. the time when an origination access attempt (seizure) is received. 2. the time when the mobile acknowledged the page (page response seizure) for terminations. 3. the time the first page attempt was sent for termination attempts that had no page response. This is ECP-based time and means that mobiles that did not acknowledge the page have a call length of zero.
4.	32 bits	Call Length	Call elapsed time The RF time of the call. This is calculated from the new call start time and the time of normal or abnormal call release or the time a hard handoff to another MSC occurs. The call start time in PCMD is taken from the traffic channel seize time. The call elapsed time is taken from the traffic channel elapsed time. If the call fails or is blocked and the traffic channel seize time is zero, then the call start
5.	6 bits	Call type	Call processing call type
6.	32 bits 56 bits	Serial number	Mobile serial number (MSN) Mobile Equipment Identifier (MEID)
7.	20 bytes	Subscriber number	Subscriber directory number
8.	10 bytes	MIN	Mobile identification number
9.	15 bytes	IMSI	International Mobile Station Identifier
10.	16 bits	Ini. Cell Site Num	Initial cell site number
11.	4 bits	Ini. Sector	Initial sector number
12.	4 bits	Initial CCU	Initial Channel Control Unit (CCU)
13.	8 bits	Ini. Channel Num	Initial channel element number
14.	16 bits	Last Cell Site Num	Last cell site number
15.	4 bits	Last Sector	Last sector number
16.	4 bits	Last CCU	Last CCU
17.	8 bits	Last Channel Element Num	Last channel element number
18.	16 bits	PSTN Trunk Group	PSTN trunk group
19.	16 bits	PSTN Member	PSTN member number
20.	11 bits	PSTN DCS	PSTN DCS number
21.	8 bits	Intra DCS Handoff	Number of intra-DCS hand-offs
22.	20 bytes	Originated Digits	Mobile originated dialled digits

23.	4 bits	Start Cell Type	Cell type at start of call
24.	4 bits	End Cell Type	Cell type at end of call
25.	6 bits	Seizure Carrier	Seizure carrier
26.	6 bits	Assigned Carrier	Assigned carrier
27.	6 bits	Ending Carrier	Ending carrier
28.	5 bits	Assigned Carrier Band	Assigned carrier band class
29.	16 bits	Received from Other MSC	Indicates if the call is an origination, a termination, or was handed in from another MSC.
30.	16 bits	Req. Service	Requested service option

TABLE 3-1. **Description of fields in the PCMD data set.** The detailed description of some relevant fields is presented here in order to demonstrate the processing complexity of my custom software that parses raw PCMD data.

3.1.2 Exemplary Traces

I present the traces used for the user-mobility, load-balancing and cell-breathing aspects of the network, in this section. For **user-mobility**, my analysis uses two distinct traces. One of the traces was collected from the Auckland area switch on 1st September, 2007. This trace is a sample of an hour-long trace and has around 65K data-samples in it. The second trace was collected from the Wellington area and is a much shorter trace, comprised of data spanning a few minutes, containing around 3K samples. The data sets used in this study, along with their sample sizes and a brief description are presented in Table 3-2.

Metric Measured	**Data-set**	**Brief Description**	**Sample Size**
User-Mobility	Abbreviated	This is an abbreviated data-set Containing some select fields from The PCMD data	2424
	Detailed_OneHour	This is an hour long trace generated on 07/09 at Auckland	65,000
Paging Characteristics			

Load-Balancing	All day data	June 7^{th}, 2008, All Blacks rugby game in Wellington	1667914
	Rush Hour	June 7^{th}, 2008, All Blacks rugby game in Wellington	127184
	Regular Weekday	March 27^{th}, 2009, Regular Weekday in Wellington	48632
	Regular Weekend	March 29^{th}, 2009, Regular Weekend in Wellington	24676
Cell-Breathing	Regular Weekday	3 July 2009, CJKL2(Jacksons Lookout) Site 2 sector, 3 carrier site	17,509

TABLE 3-2. **Traces used for measuring user-mobility, enhancing load-balancing and cell-breathing effects.** Different traces were used to substantiate these technology measurements that affect my system design directly. Contrasting traces based on length, day of the week and particularly loaded cell-sites establishes valid comparisons in my base-line measurements. Every trace is statistically significant and delineates the portions of the data that is relevant to the study.

For **load balancing** my traces (or datasets) are used in this study, including a rush hour trace and a whole day trace, collected on the day of the Rugby game, at Wellington. The data was collected in hour-long blocks and rush hour was identified as the hour in which the most calls were placed. Two additional traces were collected on March 27^{th} (a Friday) and March 29^{th} (a Sunday), 2009, in order to compare and contrast a cell-sites pilot strength with calls made at a particular distance from the cell. In the case of **Cell Breathing** I used a trace from Jackson's lookout, collected on July 3^{rd}, 2009 with around 17K samples, (henceforth referred to as CJKL2). In this case, the site has 3 sectors. Each sector can have 1 or more carriers. Calls include all origination and termination voice and data calls. SMS calls are not included in the analysis. I chose Jackson's lookout as this cell is installed near the top of Mount Taranaki and serves a significant amount of traffic over vast distances (i.e 100km). The site has a total of 2 sectors and each sector is equipped with 3 carriers. For simplicity and practicality, only sector 2 has been included in the dataset analysis. Further, the base station pilot signal's strength relative to the initial sequence offset is represented mathematically as Ec/Io where Ec (energy per chip) is the relative strength of the signal relevant to the communication and Io is the power of the entire transmission signal within the bandwidth. The unit of Ec/Io measurements is in decibels.

3.2 User Mobility

Analysis of the traces for user mobility was carried out using a combination of generating histograms and where necessary using other similar statistical tools. One particular aspect of the data analysis I wish to point out is the calculation of distance, based on latitude and longitude information. In certain cases, I needed to measure the distance traveled by the caller, when the call was carried out. While most callers stayed stationary, there was a sub-set who moved around, while talking on their cell phones. While PCMD only collects latitude and longitude of the cell site where the call started and ended, I computed the distance using the Haversine formula [9], which is presented below:

```
R = earth's radius (mean radius = 6,371km)
Δlat = lat2- lat1
Δlong = long2- long1
a = sin²(Δlat/2) +
cos(lat1).cos(lat2).sin²(Δlong/2)
c = 2.atan2(√a, √(1-a))
```

In order to analyze the characteristics, a couple of tests on how mobility is affected during the calls were performed. It is important to understand the mobility characteristics for two end goals; (a) when modelling mobile phone networks, basing the model on real-time data, the mobility aspect is the key. The model has to accurately present how the users will move within the network, (b) Other metrics such as average call-time, user-activity, user-mobility correlated with user-activity, number of cell changes etc. give further insight needed to model these networks accurately and (c) In the case of handling emergencies, understanding caller traffic is key to identifying which citizens on the ground can be contacted, what their likelihood of movement is over a period of time and how their call activity co-relates to their movement in a region. To understand how a user moves between cell-sites, when placing calls, the originating cell ID for each call in the trace is compared to the cell ID where the call ends. The number of cell site changes represents the movement of the user between calls, thereby demonstrating user mobility. Fig. 3-2 presents the user-activity observed on the longer, hour-long trace. Around 65,000 calls were monitored in this trace wherein 23% of the calls were made, while the user moved within two cell-sites. 76% of the calls were made when the users were moving more than two cell-sites, during the call. This is tracking the *total number of calls, which* are made by each unique ID.

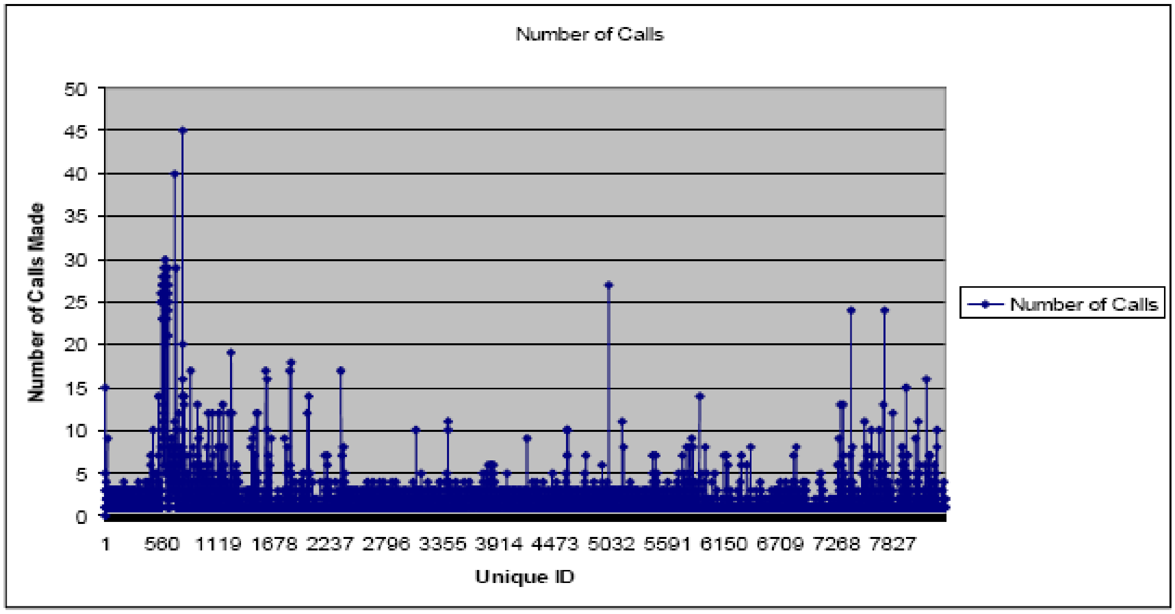


FIGURE 3-2. **User-Activity over an hour.** This graph plots the total number of calls, shown on the y-axes made by unique callers, shown on the x-axis, over an hour.

Per this plot, the number of calls placed by a *unique user* varies between 4 calls to 45, with the average number of calls being 3 and the mean and median being 1.

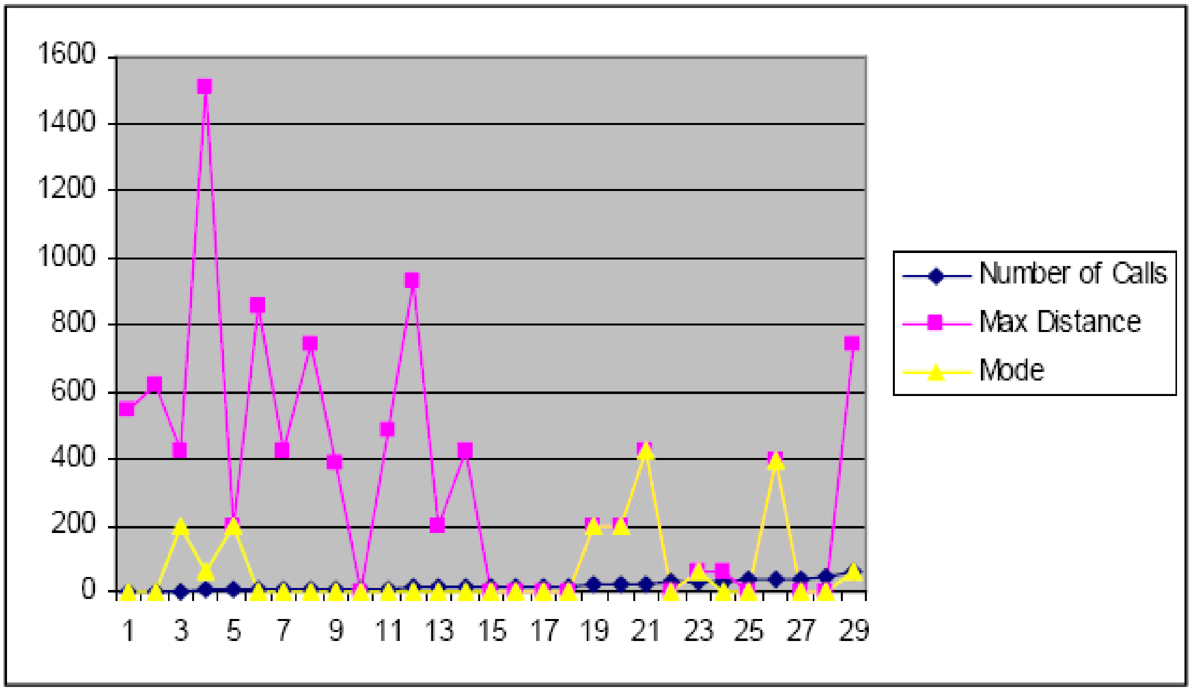


FIGURE 3-3. **User Mobility correlated with user activity level.** This graph plots the total number of calls, shown on the y-axes made by unique callers, shown on the x-axis, over an hour.

Fig. 3-3 correlates user mobility with user activity, using the shorter trace. This plot shows the number of calls made by a *unique user* who pertains to a *unique mobile phone number*. The user's activity is measured in terms of the distance travelled by the user, while the call is in progress. The user's mobility is measured by calculating the distance between the latitude, longitude tuple of the cell-site where the call was initiated and where the call ended. There are three pieces of information represented in Fig. 3-3. The Number of calls indicates just that for each unique user. The max distance indicates the maximum distance, travelled by that user, while placing his or her entire set of calls. The *mode* indicates the most popular or frequent distance, which a particular user travels, while placing their set of calls. This is important and interesting, as when modelling this data, I will need to understand what the frequently occurring distances are, in order to simulate a similar scenario for future use.

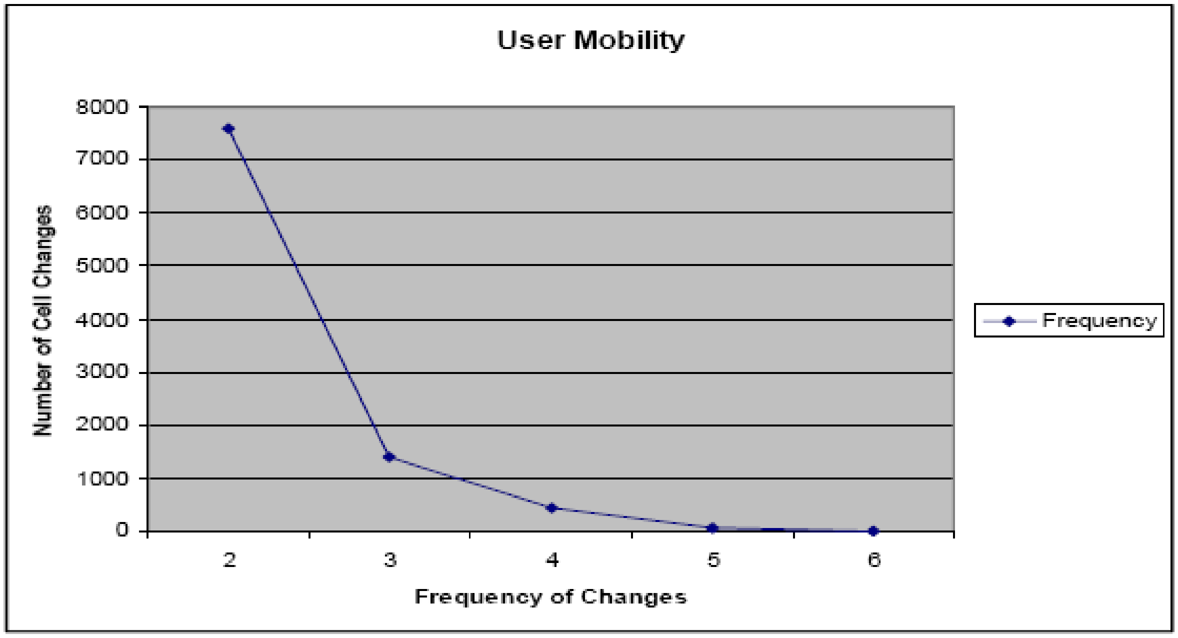


FIGURE 3-4. **User roaming range.** This graph plots user-mobility as indicated by the frequency of cell changes.

I further see that the user mobility is weakly correlated with the user's activity level. Fig. 3-4 plots the user roaming range wherein the frequency of cell-site changes is plotted against the number of cell-site changes, using 9000 unique callers, from the hour-long trace as a representative sample. This graph shows that the number of times a user moves between two cell-sites is nearly 8000 (in the entire sample of 9000 calls), which is more than 50% of the data-set. 14% of the callers or 1396 (as shown in Table 3-3) change 3 cell-sites in the duration of their call and so on.

Cell Changes	Frequency	Percentages
2	7592	54.3%
3	1396	14.7%
4	426	4.5%
5	72	0.7%
6	4	0.04%

TABLE 3-3. **Cell-Site changes.** This table shows how frequently users move between 2,3,4,5 and 6 cell sites, in the duration of a phone call. This is significant as it gives us some indication of the distances people tend to travel, on average when placing calls.

Table 3-3 summarizes the percentages of cell-site changes against frequency. Fig. 3-5 plots user-activity against user mobility wherein the number of cell-site changes are plotted against the number of calls made *by all users*, in that range of cell-site changes. 45% of the calls demonstrated a shift of about 4 cell-sites. 11% of calls are made when the user moves across 3 to 4 cell-sites. 23% of calls were made when the users moved across 2 cell-sites. Only 8.5% of the calls corresponded to users moving 6 cell-sites. This is different from Fig. 3-4, which plots the frequency of changes amongst calls made by *unique users*.

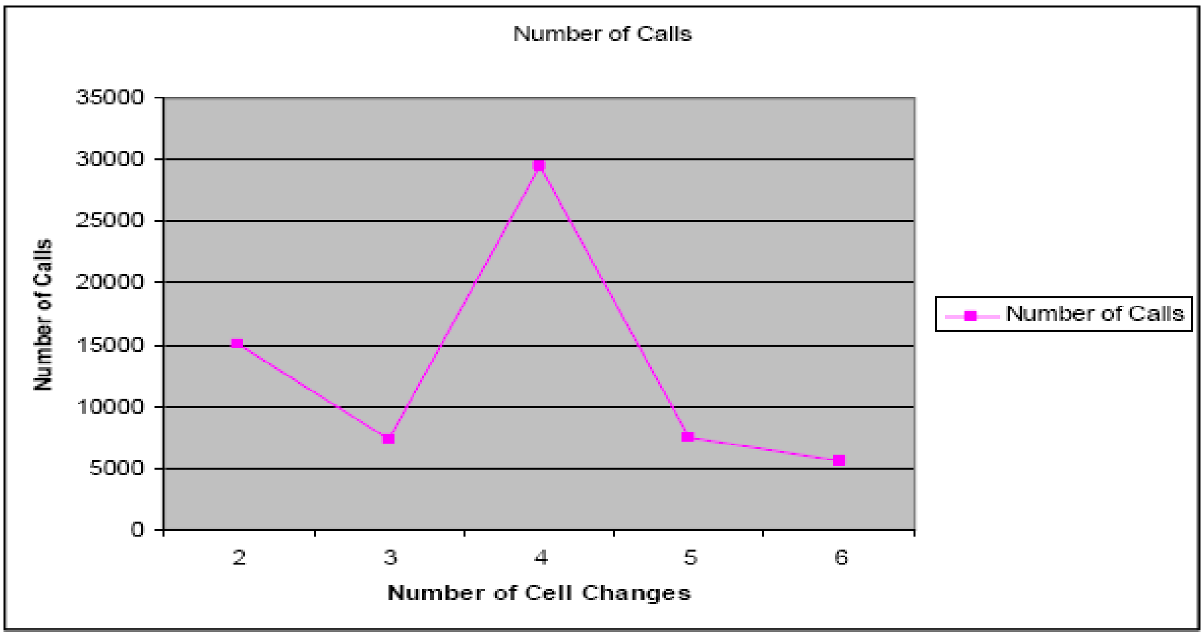


FIGURE 3-5. **User Activity vs. User Mobility.** This graph plots user-mobility as over all users.

3.3 Load Balancing

Several load-balancing techniques have been proposed and studied for CDMA systems using a variety of simulation-based data sets. In this study, I present actual call data captured at Wellington, New Zealand, from their leading telephone services provider, in order to understand how calls in the active-set are shared amongst cell-sites. Two groups of cell-sites are analyzed to understand this phenomenon including (a) The busiest cell-sites, which experience the highest call volume and (b) The cell-sites with the poorest coverage or reference pilot signal strength. I obtain a daylong trace on the date of an All Blacks

Rugby game in Wellington [40] New Zealand and further delineate the rush-hour data on this day. Rush hour is defined as the hour at which the most calls are placed during the day, which co-incided with the time at which the Rugby game started. I observe that the busiest sites are located in downtown Wellington and calls in the active set migrate in the direction of caller traffic, which was moving towards the Westpac stadium, where the Rugby game was held. Calls are not picked up by the nearest neighbours (to the busiest sites) and are absorbed by the neighbours that are proximal to the stadium, suggesting directional load balancing. In the case of cell-sites with poor coverage, the load is distributed amongst the nearest neighbours, when the coverage of a particular cell-site reduces, over time. The data used in this study was collected and segregated into three main categories of calls: calls made during rush hour, SMS and Zero length calls and calls made on cell towers with poor coverage, as measured by signal pilot strength.

SMS and Zero Length calls are self-explanatory. Every record in the PCMD data set pertains to one call and has several attributes associated with it. One of these attributes describes the kind of call placed from the phone and another attribute is the length of the call. These are profiled to filter out SMS and zero-length calls. ***Rush hour*** is defined as the hour in which the most calls are placed, during the day. As the trace contains information about the time at which the call was placed, the trace can be processed to obtain an hourly breakdown of calls. ***Calls made from towers with poor coverage*** are determined by using the Ec/Io values in combination with some service level measurements.

The base station pilot signal's strength relative to the initial sequence offset is represented mathematically as Ec/Io where Ec (energy per chip) is the relative strength of the signal relevant to the communication and Io is the power of the entire transmission signal within the bandwidth. The unit of Ec/Io measurements is in decibels. Table 3-4 describes how I split the traces apart to take into account individual characteristics of the trace. The results are averaged over all sub-traces.

Time Period	**Zero Length Calls**	**SMS Calls**	**Low pilot strength or coverage**
	No	No	Yes

All day data for June 7th, 2008	No	Yes	Yes
	No	Yes	No
	No	No	No
Rush Hour Data for June 7th, 2008	No	No	Yes
	No	Yes	Yes
	No	Yes	No
	No	No	No

TABLE 3-4. **Experiments performed.** This table shows how frequently users move between 2,3,4,5 and 6 cell sites, in the duration of a phone call. This is significant as it gives us some indication of the distances people tend to travel, on average when placing calls.

In this section I seek to characterize load-balancing effects, as observed over the geographical area of Wellington, New Zealand. Cell sites are monitored based on a unique numeric identifier, associated with every cell site. The number of calls placed at a cell site is collected and stored for analysis. This information coupled with other metrics pertaining to the pilot strength of the reference pilot, and the distances at which calls in the active set are placed are taken into consideration, in order to characterize the data properly.

3.3.1 Cell Loading statistics based on call volume and distance

In this section, I discuss the cell loading statistics based on number of calls made at a cell site, the pilot strength of that particular site and the distance from which the calls in the active set are placed.

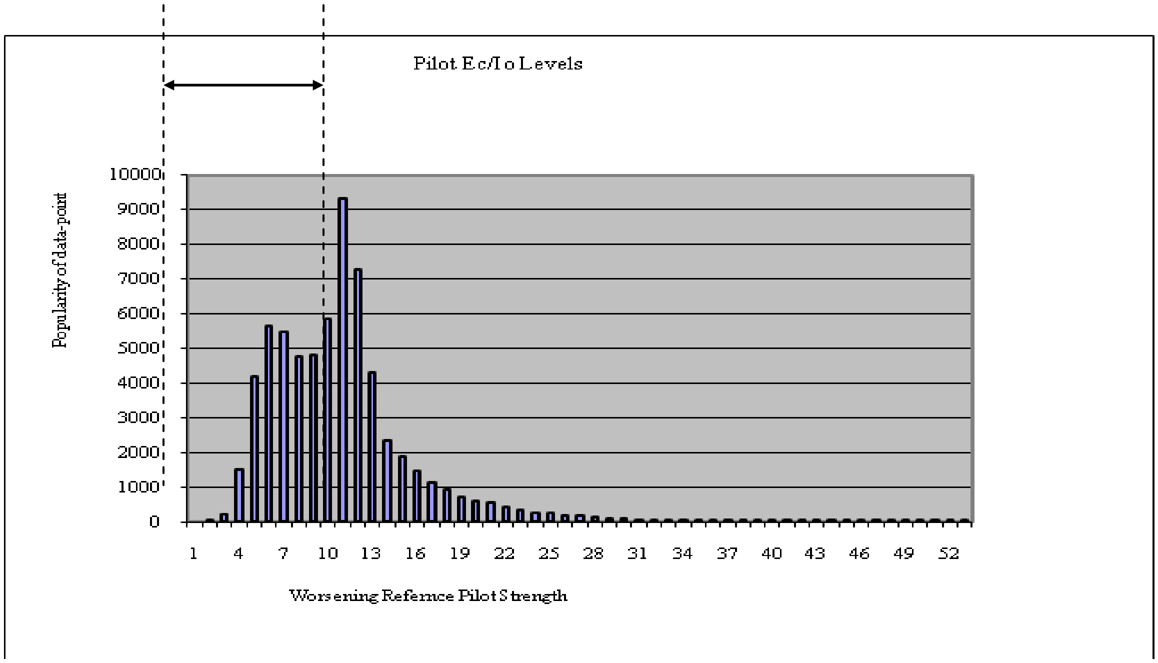


FIGURE 3-6. **Pilot Ec/Io Levels.** Number of calls placed at a particular cell site vs. the pilot strength on the reference pilot for that site.

Fig. 3-6 shows the number of calls made across all cell sites in a data set that excluded zero-length calls over the daylong trace. The x-axis shows the pilot strength (or normalized Ec/Io values, remember they are negative values and therefore need to be normalized) of the reference pilot as it weakens from left to right and the y-axis shows the number of calls placed. It is observed that on pilots with poor coverage, the total number of calls made drops off. This is because although a call may be placed at a particular site, if the coverage is too poor, the call fails. As the coverage of a site deteriorates, the number of calls made drops because the site is unable to allow the call to go through.

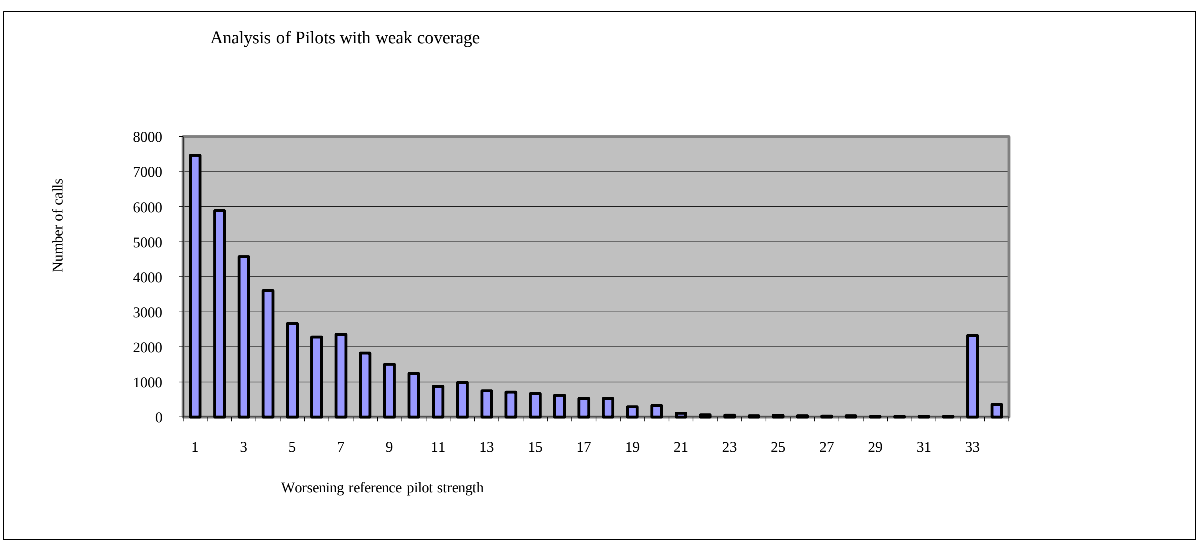


FIGURE 3-7. **Analysis of pilots with weak coverage.** Call volume at cell sites with poor coverage.

Fig. 3-7 places in focus the calls made on cell sites with poor coverage in a data set that excluded zero-length calls over the daylong trace. The x-axis shows the pilot strength (or normalized Ec/Io values) of the reference pilot as it worsens from left to right and the y-axis shows the number of calls placed. It is seen in this figure that there are a few very cell sites with poor Ec/Io values, wherein the number of calls placed is still high. This could be explained as reflecting the calls that possibly failed to originate or those that terminated abnormally. There are an anomalously high number of calls at the right-hand side of the graph at a very poor coverage value and I point out that these are probably calls that terminated abnormally.

Since the data was collected on the date of an All Blacks rugby game, two candidate sets of cell sites can be scrutinized to understand how calls in the active set migrate amongst cells from which they originate and their immediate neighbors: (a) Monitoring cell towers in downtown Wellington, close to the stadium where the rugby game was being played and (b) Monitoring cell sites with extremely poor coverage. As downtown experiences a vast spike in call traffic, as the hour of the game approaches I am able to study the effects of how calls in the active set are shared and infer whether it is the closest neighbors that pick up the calls from congested sites or neighbors more proximal to where majority of the traffic is moving.

The distance from which calls are made is defined as the distance at which the caller is located from the base of the cell site and this is calculated according to equation 1, by using the round-trip delay time, which is calculated on a per-call basis.

```
RTD(chip): [CDMA_RT_DELAY]/8

Access Dist (km): [CDMA_RT_DELAY]/16*0.243
```

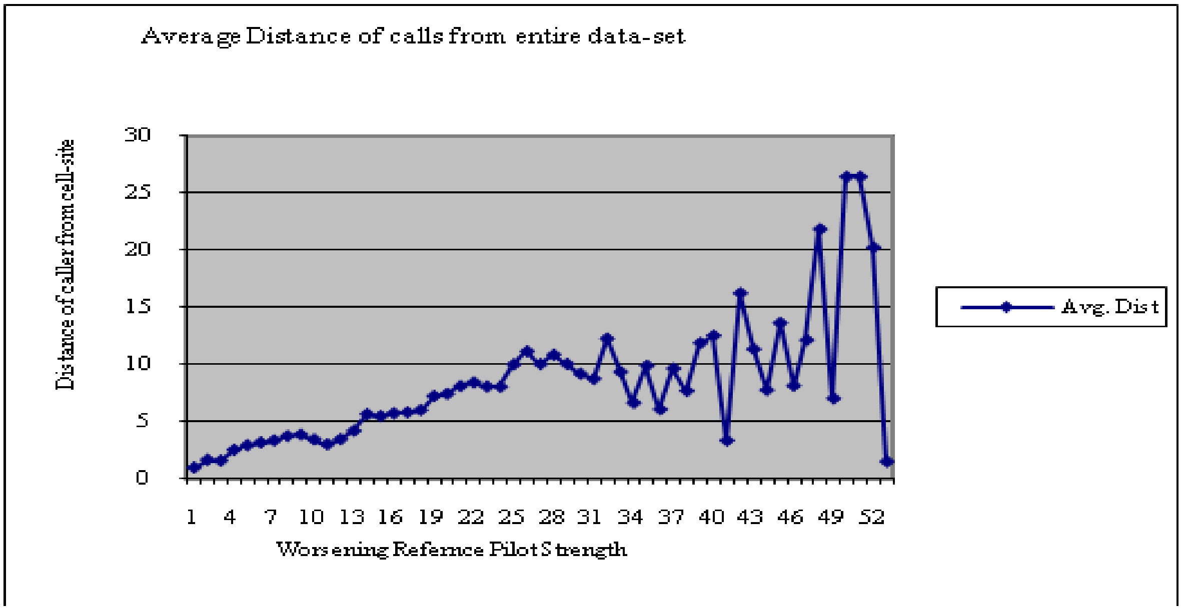


FIGURE 3-8. **Average Caller Distance.** This graph plots the average distance of calls placed within cell sites in the daylong trace, for all cell-sites.

Fig. 3-8 shows the average distance at which calls were placed on all cell towers in the daylong trace. The x-axis shows the pilot strength (or normalized Ec/Io values) and the y-axis shows the distance from which the calls are placed. I see that with poorer coverage, the average distance from which the calls are placed seems to increase. Fig. 3-9 focuses on those cell-sites with coverage worse than the threshold and shows the average distance at which calls were placed on cell towers, with increasingly poor Ec/Io values, in a data set that excluded zero-length calls over the daylong trace. The x-axis shows the pilot strength (or normalized Ec/Io values) and the y-axis shows the average distance from which calls were placed. At the worst pilot strength values, the distance at which calls placed drops suddenly and this anomaly is once again explained as calls that failed to originate or were terminated abnormally.

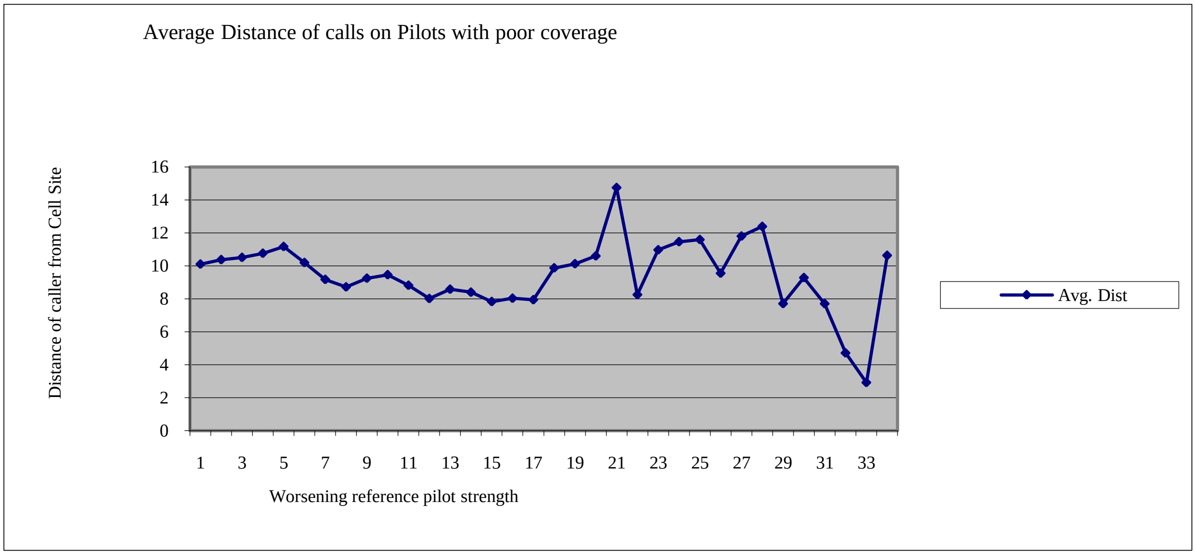


FIGURE 3-9. **Average Caller Distance with worsening pilot strength.** This graph places emphasis on cell-sites where coverage is worse than threshold and relates this back to the distance at which the calls were placed.

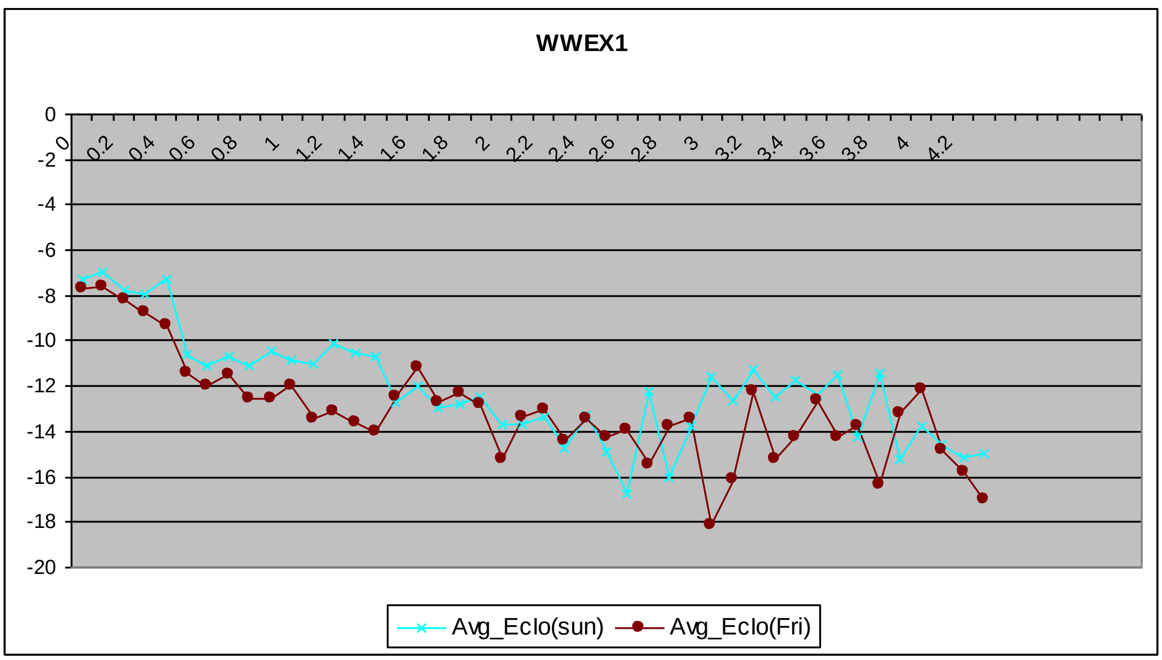


FIGURE 3-10. **Average Caller Distance on Friday vs. Sunday.** This graph shows average distance of calls placed within one particular cell-site on a Friday and a Sunday to compare Ec/Io values at similar distances.

Fig. 3-10 plots the average distance of the calls placed on a Friday (March 27th) as opposed to a Sunday (March 29th) when the call volume is about 50% that of the call volume on Friday (Table 1). The x-axis shows the distance in kilometres and the y-axis plots the unnormalized Ec/Io. I observe that the Ec/Io values beyond 3km appear to be Weaker (when compared between Sunday and Friday). This could possibly due the users at the edge of the cells experiencing poorer Ec/Io due to the increased number of calls made on the cell. This would create a good candidate set of cells that can be analyzed to understand how calls are shared.

3.3.2 Busy Cell Sites demonstrating load balancing

Table 3-5 shows the number of calls placed from the busiest cell sites, which have unique numerical identifiers associated with them. The ten busiest sites are collected in the all day trace and the rush hour trace in order to compare and contrast the load on the sites. These are later juxtaposed with the Ec/Io values on these cell sites in order to understand whether the cells drop calls or shed their load to their neighbors. Neighboring cells are often located at various radii from the originating cell and absorb calls as the coverage or signal strength of the originating cell site fluctuates.

Cell Descriptor	Cell-Site Number	Number of calls placed
All Day Data	219	36762
	348	36565
	222	35141
	99	34097
	66	33313
	100	32883
	168	32637
	186	32365
	40	31487
	12	30819
Rush-Hour	186	2768
	66	1916
	222	1820
	219	1787
	100	1738
	348	1701
	99	1634
	96	1618
	12	1571
	228	1564

TABLE 3-5. **Unique Cell Site Statistics.** This table shows the number of calls made from unique cell sites. The Unique ID of the cell-site is shown in Column 1 and the number of calls in Column 2. I have shown the cell sites sorted by the most traffic in order to discuss the most relevant sites, pertaining to load-balancing.

As seen in this table, cell site 186 is one of the busiest sites during rush hour, which I know to be the time before the Rugby game, as the traffic congregates at downtown

Wellington. Picking this cell site is a valid choice as even during the day, this cell site receives a large number of calls. The traces used to identify the busiest sites were compared across those in which the zero-length calls and the SMS data was included and excluded and 186 are established to be the busiest site.

Fig. 3-11 shows the relative placement of cell-site 186 and its immediate neighbours that have also been observed to handle a high volume of calls, in order to understand **load** balancing effects with calls in the active set. The distance of cell site 186 from its neighbours has also been indicated in the figure. The nearest neighbour is 208 at 0.7 KM followed by site 60 at 0.8KM, site 178 at 2.5KM, site 298 at 2.6KM and site 184 at 2.7K.

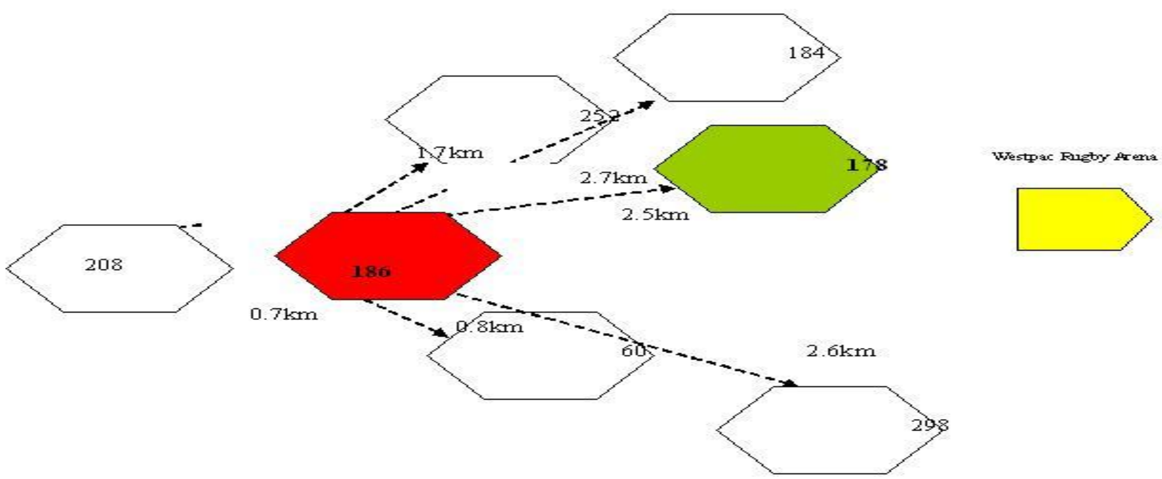


FIGURE 3-11. **Cell #186 and its neighbors.** The distance between cell-site 186 and its immediate neighbors is shown in this figure, as is their relative position to the Westpac Rugby Arena (shown in Yellow), to orient the reader towards how things are situated.

I notice the relative distances between the busiest site and its immediate neighbours as I expect that when cell site 186 sheds its load (or shrinks) it will do so to its immediate neighbours.

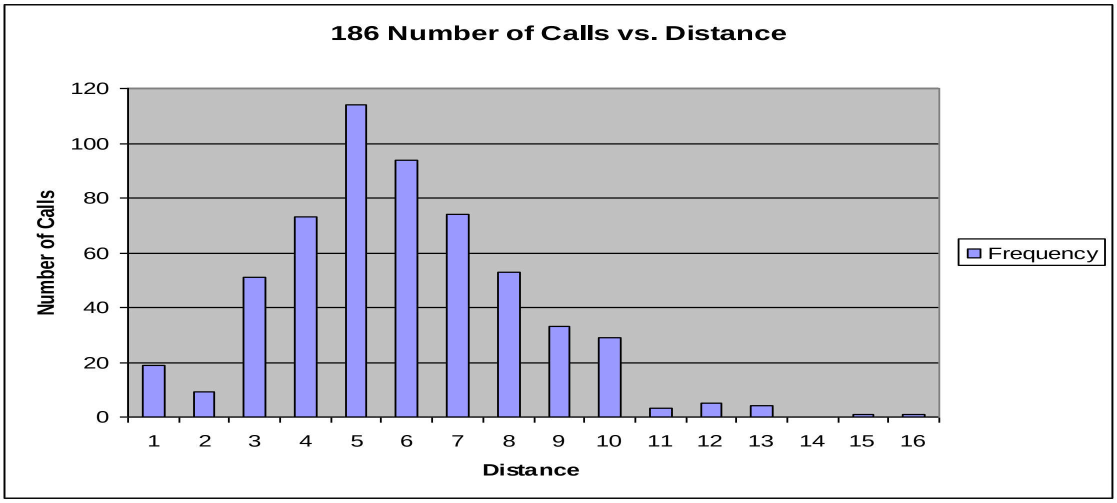


FIGURE 3-12. **Call traffic on Cell #186** . Average number of calls placed vs. distance of calls placed within cell site 186, the busiest site during rush hour.

Fig. 3-12 shows the number of calls made to cell site 186, vs. the distance (shown in km) from which the calls are placed. The x-axis shows the distance from cell-site 186 and the y-axis shows the number of calls that were placed from that distance. I see that the average distances from which calls are placed are 5km-6km. These calls are in the active set of the most loaded site in the trace as this is not an arbitrary cell site that absorbs calls from its periphery or from its neighbors. Fig. 3-13 shows the number of calls made to cell site 178, (which has the highest number of calls after site 186, compared to the rest of the neighbors and compared to site 252 which is the nearest neighbour by distance but is not as busy as site 178), vs. the distance from which the calls are placed.

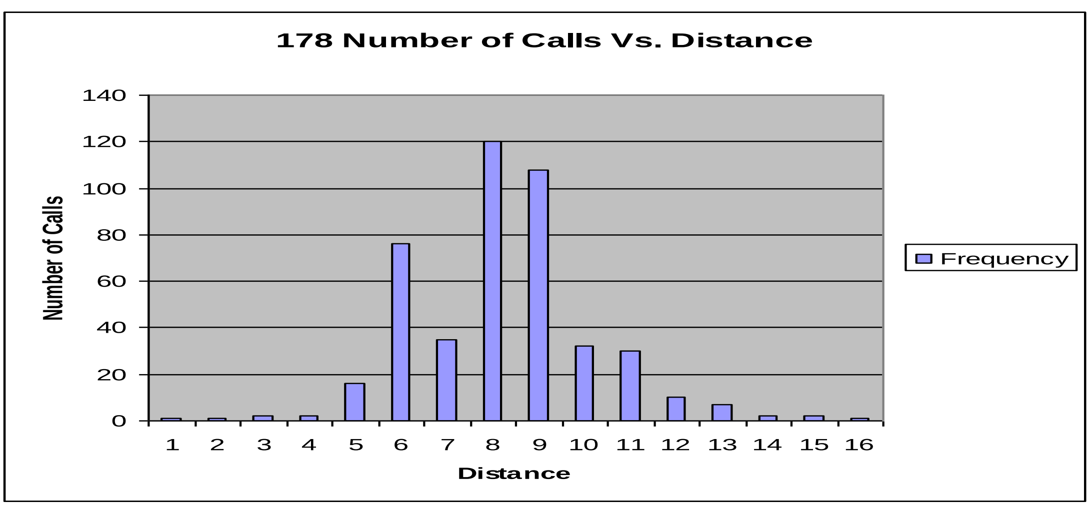


FIGURE 3-13. **Call traffic on Cell #178.** Average number of calls placed vs. distance of calls placed within cell site 186, the busiest site during rush hour.

The x-axis shows the distance from cell-site 178 and the y-axis shows the number of calls that were placed from that distance. I see that the average distances from which calls are placed are 8km-9km. This implies that site 178 is picking up the load from the periphery of site 186 in order to avoid calls being dropped.

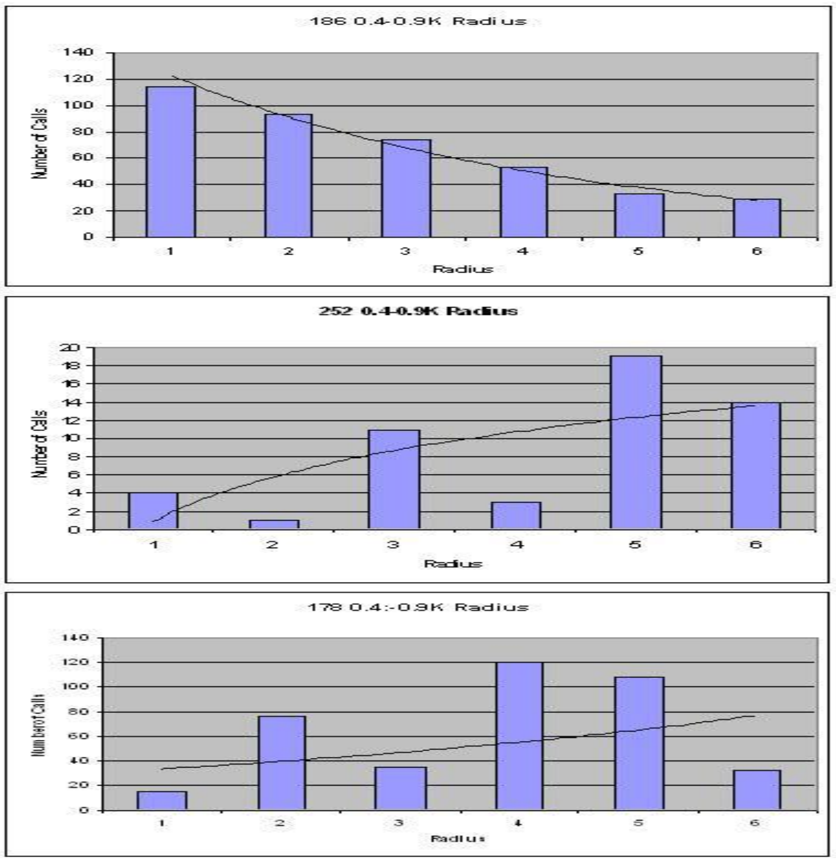


FIGURE 3-14. **Call traffic neighboring cells. Absolute** and exponential rendition of the number of calls observed at a 0.4km-0.9km radius around cell site 186 and its immediate neighbors at rush hour.

Fig. 3-14 shows the exponential and absolute number of calls made from a radius of 0.4km-0.9km on the busiest cell site 186 and its two neighbours, 252 and 178. I see that 178, which is closer to Westpac Arena where the Rugby match is held picks up a larger portion of the load shed by cell site 186. These graphs are a snapshot at one instance in time and I would expect the closest neighbour, cell 252, which is placed closer to the busiest site to pick up a majority of the calls shed by 186. However, what I observe is that it is in fact cell 178, which is closer to the Westpac stadium that is observing most of these calls.

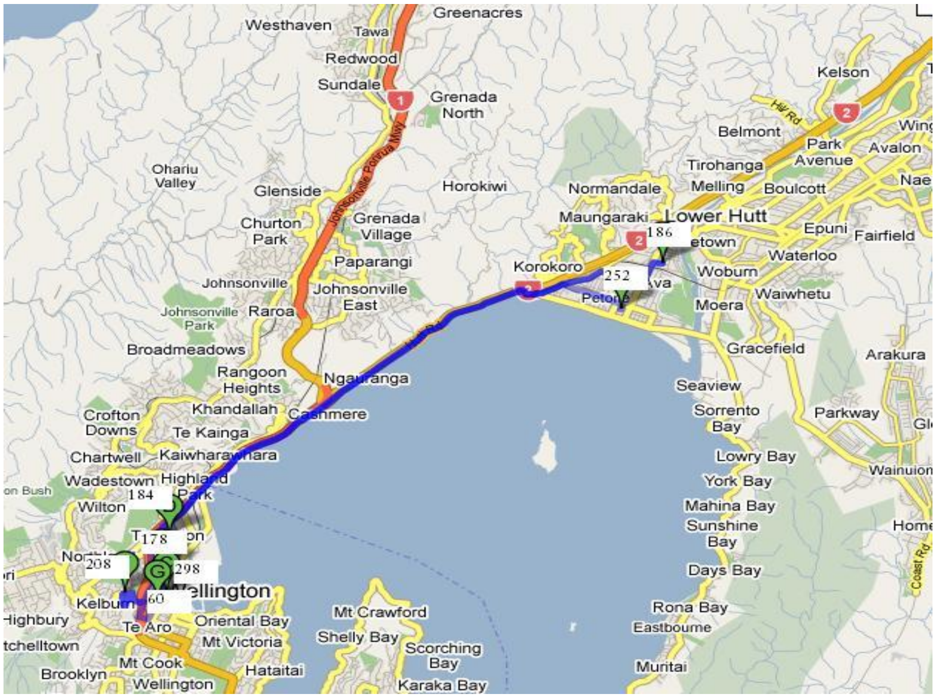


FIGURE 3-15. **Map-View of neighboring cells. Geographic** location of the seven cell sites (186 and neighbors) on a map of Wellington

Fig. 3-15 shows the geographic location of the seven cell sites (186 and neighbors) on a map of Wellington. Table IV shows the location of the cell-site alongside the street address of the site and the distance of the neighbors from site 186. The Westpac arena, where the Rugby match was held on June 7th, 2008, whose rush hour call traffic is being analyzed in this section, is located at Waterloo Quay in Wellington. This location is 2.5km from cell-site 186 (Wellington Central), 0.4km from cell-site 178 (Lambton Quay) and 0.8km from cell-site 252 (Bolton Quay). The proximity of the cell-site to the stadium explains why cell site 178 picks up more of the load shed from 186 (even though site 252 is CLOSER to site 186, when compared to 178). As the traffic moves towards the stadium, the sites en-route picks up calls.

Cell-Site Number	**Geographic Location**	**Distance from 186 in km**
186	Central Wellington	0
252	Bolton Quay	1.7
178	Lambton Quay	2.5
298	Victoria Villeston	2.6
184	Thorndon	2.7
208	Kelburn	0.7
60	Majestic Center	0.8

TABLE 3-6. **Street Location of Neighbouring Cell Sites.** The seven neighbouring sites shown here are approximately located at Central Wellington, Bolton Quay, Lambton Quat, Victoria Villeston, Thomdon, Kelburn and Majestic Center. These are approximate locates as the data and locations have been anonymized.

3.3.3 Cell Sites with poor coverage demonstrating load balancing

In this section, I briefly analyze the load balancing amongst the cells with the poorest coverage (or worst Ec/Io values) in my daylong trace collected at the Wellington switch. Since this switch collects data for a broader geographic area than just Wellington, the cells with poor coverage was identified to be cell-site 96, Blagdon Hill, whose six immediate neighbors were cell sites 99, 93, 261, 343, 262 and 260. The location of the site with the poorest coverage and its immediate neighbors and distance form site 96 is presented in Table 3-7.

Cell-Site Number	**Geographic Location**	**Distance from 186 in km**
96	Blagdon Hill	0
99	New Plymouth City	3.7
93	Brooklands	5.7
261	Fitzroy	6.9
343	Mangorei	9.1
262	Bell Block	11
200	Oakura	13

TABLE 3-7. **Street Location of Neighbouring Cell Sites with poor coverage.** The seven neighbouring sites shown here are approximately located at New Plymouth, Blagdon Hill, Brooklands, Fitzroy, Mangorei, Bell Block and Oakura. These are approximate locates as the data and locations have been anonymized.

The absolute distances are seen to be a lot farther amongst neighbours in this area as this is not a bustling city. Fig. 3-16 shows the location of the three neighbours, with respect to each other. Fig. 3-17 shows how the Ec/Io levels vary between site 96 (Blagdon hill) and its immediate neighbours 93 (Brooklands) and 99 (New Plymouth City). As seen in this figure, the closest cell site absorbs most of the load when tower 96 starts to drop its calls. The absorption of calls by the neighbours of cell site 96 does not have a directional component to it, as was the case with the busiest site 186 where the calls were being absorbed in the direction of movement of caller traffic. As these traces were captured with time-stamps on each of these calls, the x-axis shows the time of the call on each site and the y-axis shows the number of calls made at that instant in time, on that particular site for the three sites.

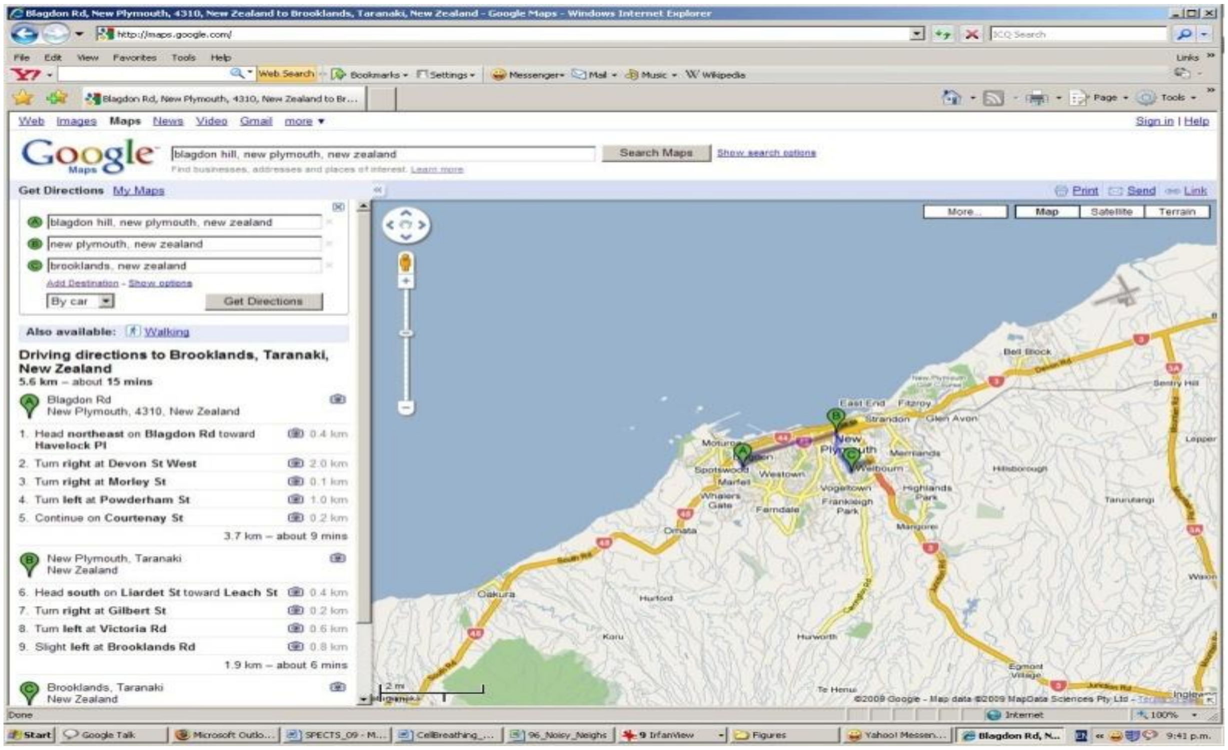


FIGURE 3-16. **Map-View of neighboring cells** .Location of the tower with poorest coverage and it’s two immediate neighbors.

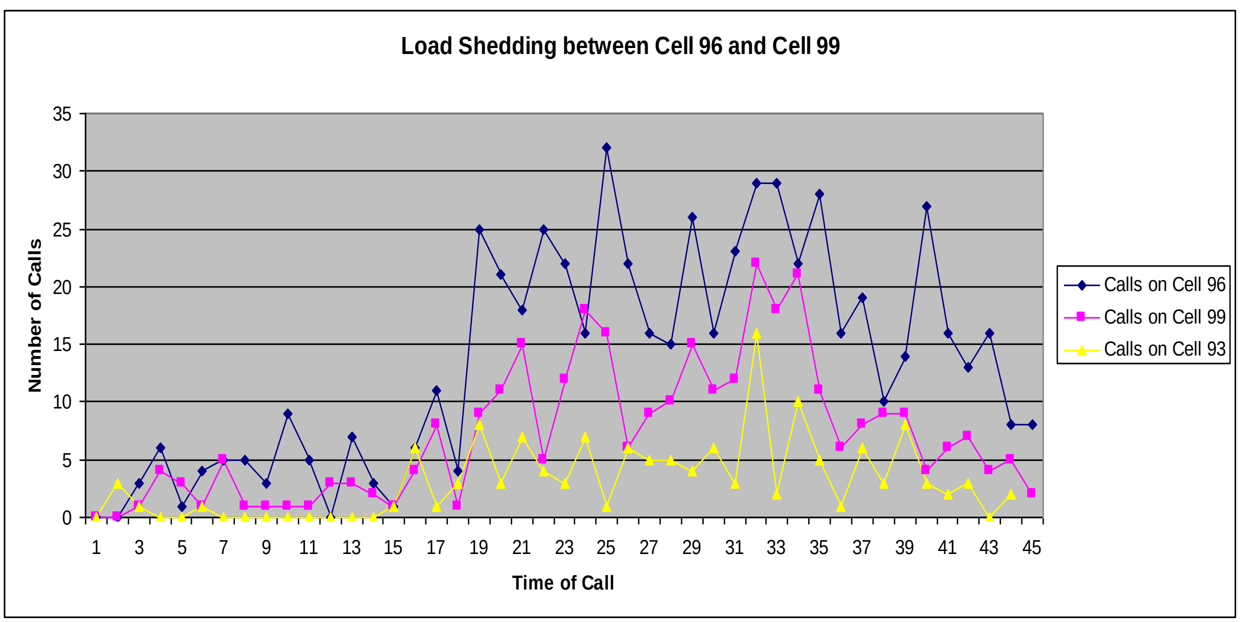


FIGURE 3-17. **Load Shedding across time** . Load balancing amongst cell sites with poor coverage.

3.4 Cell Breathing

Cell Breathing in CDMA networks has been studied in order to perform better congestion control and enable proper planning of network resources. Previous studies that have been performed on cell breathing effects have focused on a simulation-based approach or have utilized data from various performance counters. In my study, I use Per Call Measurement Data (PCMD) to study the effects of Cell Breathing in a CDMA network within New Zealand. PCMD is a more detailed and accurate data-set whose granularity is per call, placed on the network. This data is collected from three switches within New Zealand, belonging to the country's leading telephonic services provider. The objectives of this work are to understand cell-breathing effects by analysing call-volume at the site over the course of the day, isolating the effects of distance from the cell-center at which the call was placed, understanding the effects at the cell-edge and presenting the effects on the call-traffic. Per call measurement data was collected at one anonymized site within New Zealand, on July 3rd, 2009, in order to analyze and present the effects outlined above. This thesis presents preliminary results that substantiate the effects of cell-breathing, by means of the PCMD data, quantify the impact of cell breathing at the cell edge by analysing increase in noise, dropped calls and lost calls, etc. All mobiles in a CDMA system use the same carrier frequency at the same time. This effectively generates interference between users in the same cell and in neighboring cells. The CDMA spreading and dispreading process allows some degree of suppression to this 'interference'. The level of suppression is termed 'processing gain'

Cell breathing (reverse link) The BS noise floor on a cell rises as loading on the cell increases. Mobiles must therefore transmit additional power to overcome the increased interference. As a result, the tolerance within the cell to overcome the maximum path loss is decreased. The cell effectively 'shrinks'. This is effectively cell breathing on the 'reverse link' (mobile→base station). In most scenarios, the reverse link is the limiting factor (max transmit power) in relation to coverage matching. ROC (Reverse overload control) is introduced in CDMA systems to estimate loading and performs the function of load management. Reverse link overload control aims to keep the reverse link loading at an acceptable level in order to protect the integrity of the reverse link. Reverse link loading in a particular sector of a cell is determined from the total interference power from mobiles operating in that sector and from mobiles operating in adjacent sectors. As the number of users increases, the total power received at the base station correspondingly increases. The ratio of the median of the total received power to the background noise power is defined as the median rise. Figure 3-18 illustrates the median rise, expressed in dB, as a function of reverse link loading, expressed in percent of theoretical capacity. (i.e the 5dB noise rise is normally selected as the engineering guide, 75% loading). The median rise is a non-linear function of reverse link loading. It increases dramatically as the loading approaches the theoretical (pole) capacity. The steep median rise under high loading conditions can cause system performance degradation.

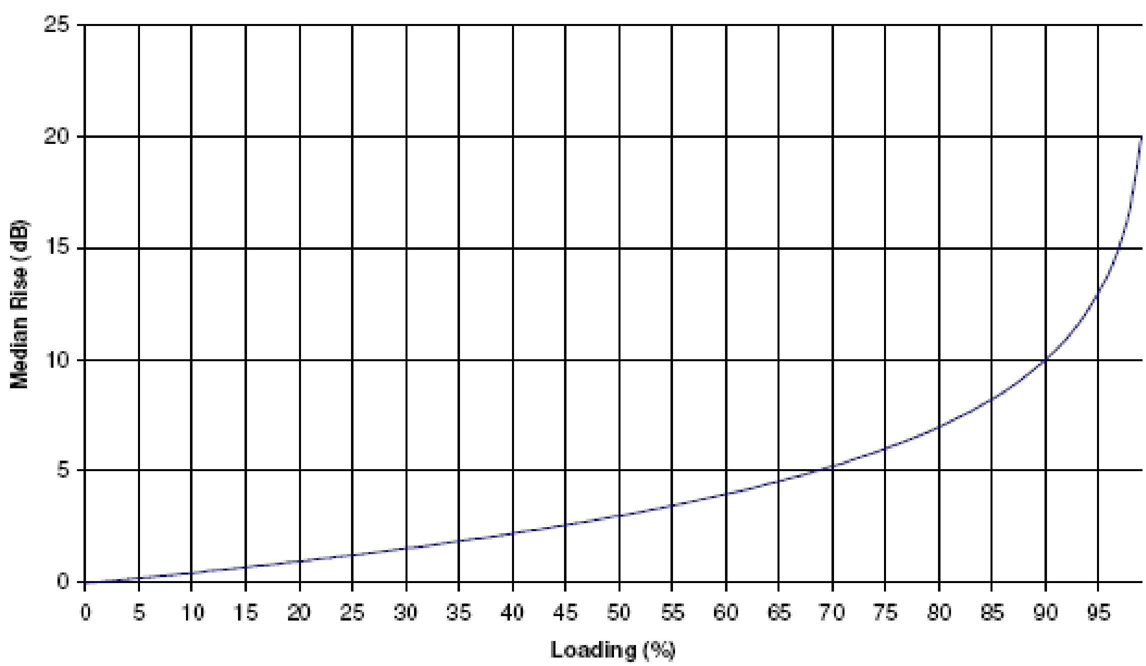


FIGURE 3-18. **Link Loading** . Median rise as a function of reverse link loading

Cell breathing (forward link) For the forward link, in a CDMA system, the total transmit power in the forward link is proportional to the number of active forward link channels weighted by the square of the respective digital gains. The total transmit power increases as the number of users increase. Control of the total transmit power is necessary for two reasons. The first reason is that the total average power that an amplifier can provide, or the amount of**.** LAC power allocated to a CDMA system, places an upper limit on the total transmit power. The transmit amplifier can be overdriven if the total transmit power is not restrained. The spurious emissions caused by overdriving the amplifier can cause interference to other services and violate FCC emissions requirements. The second reason is that the performance of the CDMA system depends on the pilot Ec/Io. As the number of users increases, the total transmit power increases and the pilot Ec/Io decreases for each user since the power transmitted for each user contributes interference to the other users served in the sector. Forward link overload control is required to prevent an increase in the total transmit power that would degrade the pilot Ec/Io beyond acceptable limits.

In this study, a one-day trace was collected at Jackson's Lookout, which is a 2 sector, 3 carrier site. Calls included in the trace contain originating and terminating voice and data calls. SMS calls are not included in this data set. This data was collected on Friday, 3rd of July, 2009 and has around 18,000 calls profiled, over the course of a 24 hour period and is summarized in Table 3-8. I characterize cell breathing based on total call volume; effects on call performance (by recording calls dropped) and user-mobility within the cell. Indicator of dropped calls include lost calls as Ill. The mobility of users across the network has been analyzed in earlier work. The sites were selected based on a busy day in the Week (Friday) number of calls made and the extend of coverage a cell serves. Most of these sites that meet this criteria are rural sites I chose Jackson's lookout as this cell is installed near the top of Mount Taranaki and serves a significant amount of traffic over vast distances (i.e 100km).For simplicity, only sector 2 has been included in the dataset analysis. This allows us to better observe any significant performance changes on the cell edge and is depicted in Fig. 3-19.

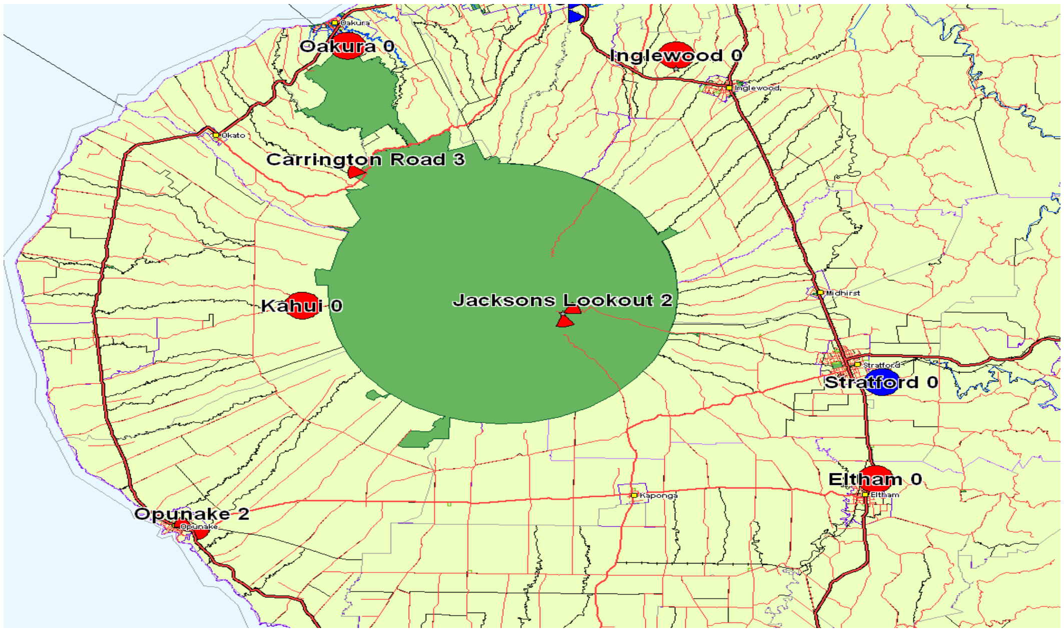


FIGURE 3-19. **Jackson's Lookout**. Location of Jackson's lookout on top of Mount Taranaki

3.4.1 Indicators of Cell Breathing

In this section I seek to characterize cell breathing effects, as observed over the geographical area of Mount Taranaki, New Zealand. Cell breathing occurs when a cell is overloaded with calls and as a result, there is an increase in interference from all the other users. The mobile must transmit additional power to overcome this increased interference. To limit this interference, the base station's reverse link overload control algorithms is used to control the reserve link loading (by way of blocking new users). Therefore, the most common cause of cell breathing is due to limitation on the reserve link.

In addition, the coverage of the site as represented by its Ec/Io value also degrades as more users make calls on the same cell site. As the number of users increases, the total transmit power of the base station increases and the pilot Ec/Io decreases for each user since the power transmitted for each user contributes interference to the other users served in the sector. Forward link overload control is required to prevent an increase in the total transmit power that would degrade the pilot Ec/Io beyond acceptable limits. This represents the second cause of cell breathing due to the limitation on the forward link.

3.4.2 Call Volume

To observe the effects of cell breathing on CJKL2, the distribution of call volume (represented as a percentage to total calls on that hour) is shown on hours 7am (when traffic is relatively light) , 11am(when call attempts are at its highest) and 11pm (when traffic becomes light again).

DATA_HOUR	Call Attempts
1	369
2	333
3	349
4	348
5	333
6	375
7	457
8	627
9	930
10	1037
11	1242
12	1013
13	1078
14	1114
15	1016
16	985
17	905
18	847
19	829
20	857
21	732
22	625
23	580
24	505

TABLE 3-8. **Call volume.** Total number of calls made at different hours of the day.

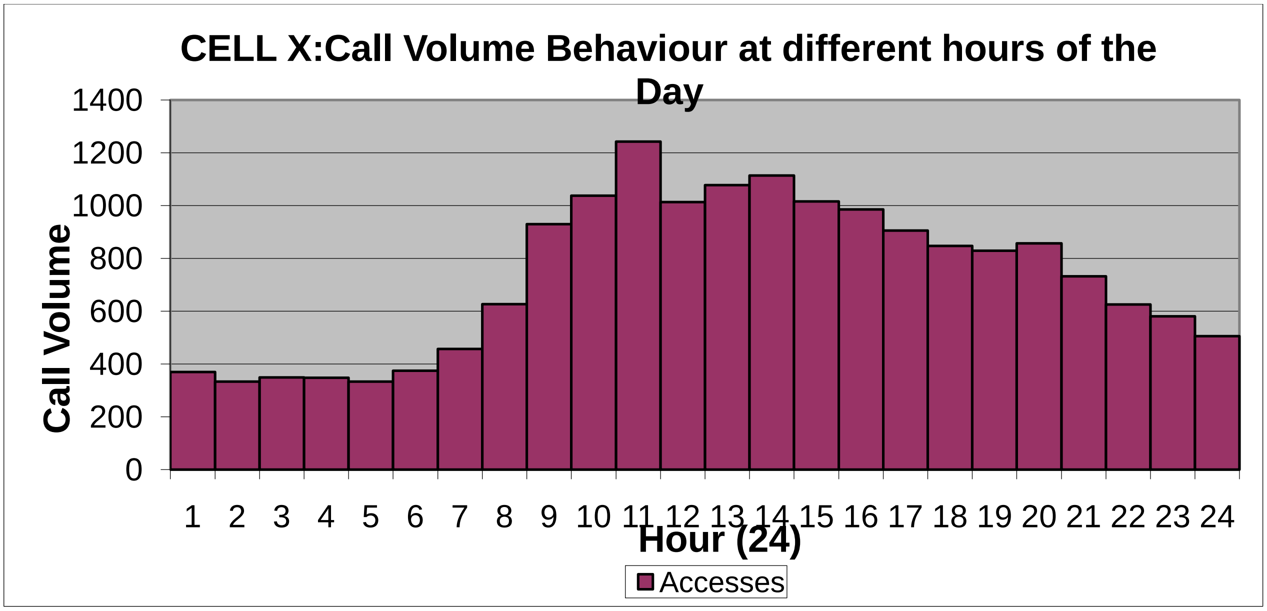


FIGURE 3-20. **Call Volume**. Call Volume behavior at different hours of the day

To observe the effects of cell breathing on CJKL2, the distribution of call volume (represented as a percentage to total calls on that hour) is shown on hours 7am (when traffic is relatively light) , 11am(when call attempts are at its highest) and 11pm (when traffic becomes light again). Table 3-8 and Fig. 3-20 show this.

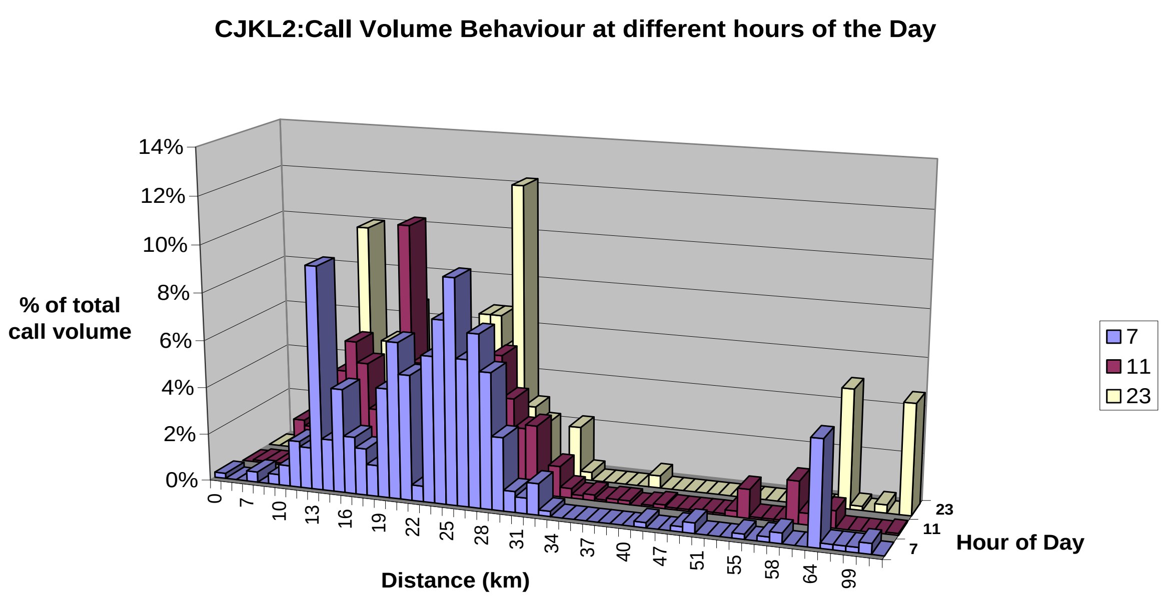


FIGURE 3-21. **Call Volume at 7am, 11am and 11pm**. Percentage of calls made at different hours of the day when the network is lightly or heavily loaded.

Fig 3-21 clearly shows that when traffic is light (7am), users are able to make calls at the cell edge (>60km). However, call attempts at cell edge are significantly reduced when the cell begins to 'breathe' (11am), resulting in a smaller cell coverage footprint. When traffic becomes light again (11pm), users are able to make calls again past 60kms.

3.4.3 Call Performance

Call performance in mobile networks is most commonly quantified via the means of call drops (including lost calls) and call blocking. In relation to cell breathing, PCMD will not likely record all types of call blocking as the call might not even register on the network due to lack of coverage or when the network is deaf to the mobile (maximum transmit power of mobile reached). As such I will only present my finding in regards to call drops that are recorded.

Hour	**Drops**	**Establish Calls**	**Blocking Rate**	**Drop Rate**	**Call**
1	14	367	99.46%	96.19%	
2	3	334	100.30%	99.10%	
3	9	348	99.71%	97.41%	
4	10	343	98.56%	97.08%	
5	5	327	98.20%	98.47%	
6	19	376	100.27%	94.95%	
7	22	452	98.91%	95.13%	
8	35	600	95.69%	94.17%	
9	56	915	98.39%	93.88%	
10	33	1069	103.09%	96.91%	
11	54	1204	96.94%	95.51%	
12	44	993	98.03%	95.57%	
13	54	1060	98.33%	94.91%	
14	53	1105	99.19%	95.20%	
15	52	1007	99.11%	94.84%	
16	35	971	98.58%	96.40%	
17	36	896	99.01%	95.98%	
18	21	841	99.29%	97.50%	
19	46	805	97.10%	94.29%	
20	26	839	97.90%	96.90%	
21	40	698	95.36%	94.27%	
22	40	600	96.00%	93.33%	

23	16	554	95.52%	97.11%
24	30	490	97.03%	93.88%

TABLE 3-9. **Call Performance.** Calls established, dropped and percentages of calls blocked and dropped are shown in this table for a 24 hour period.

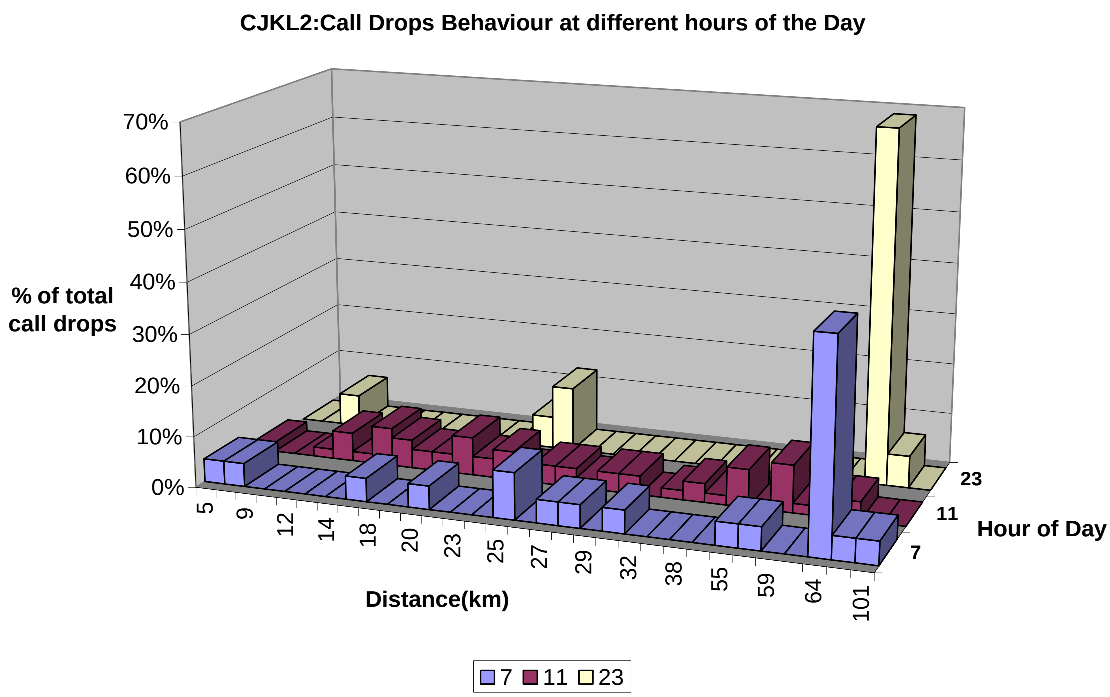


FIGURE 3-22. **Call drop distribution**. Percentage of calls made at different hours of the day when the network is lightly or heavily loaded.

The drop call distribution that occurs can be seen in Fig 3-22. Call drops appear to be at its highest when the network is lightly loaded. This phenomenon is likely occurring because the coverage area of the cell has likely reached areas (~64km) where there are no handoff with other sites. Another possibility is that the mobile's transmit power has reached its maximum, resulting in an unstable RF link which is frequently torn town as the network cannot heard transmission coming from the mobile. In fact, during the busy hour (11am), when cell breathing occurs, the drops on the network is less as the mobile users are unable to get coverage at cell edge.

3.4.4 Interference and coverage

Fig 3-23 shows the cell breathing impact more clearly over the entire day on total calls made > 60km. The percentage of calls made >60km begins to reduce to ~1.5% during the day and increases again after at night.

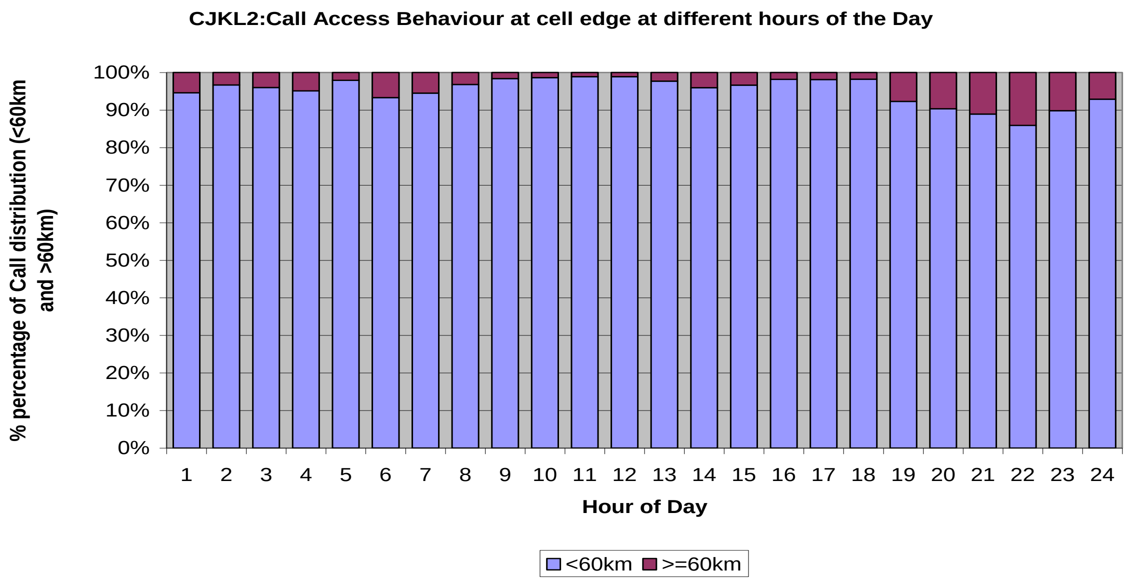


FIGURE 3-23.**Performance at 60K.** This figure shows the performance of the cell at 60ks, which is established as a significant point in the coverage area of the cell.

This is indicated by the Ec/Io value where Ec (energy per chip) is the relative strength of the signal relevant to the communication and Io is the power of the entire transmission signal within the bandwidth, is one metric that could indicate that cell breathing is occurring. The unit of Ec/Io measurements is in decibels. Each mobile user will report its first and last seen Ec/Io in the PCMD data call record. The reported Ec/Io (indicative of CJKL forward link coverage) can seen in Fig 3-24.

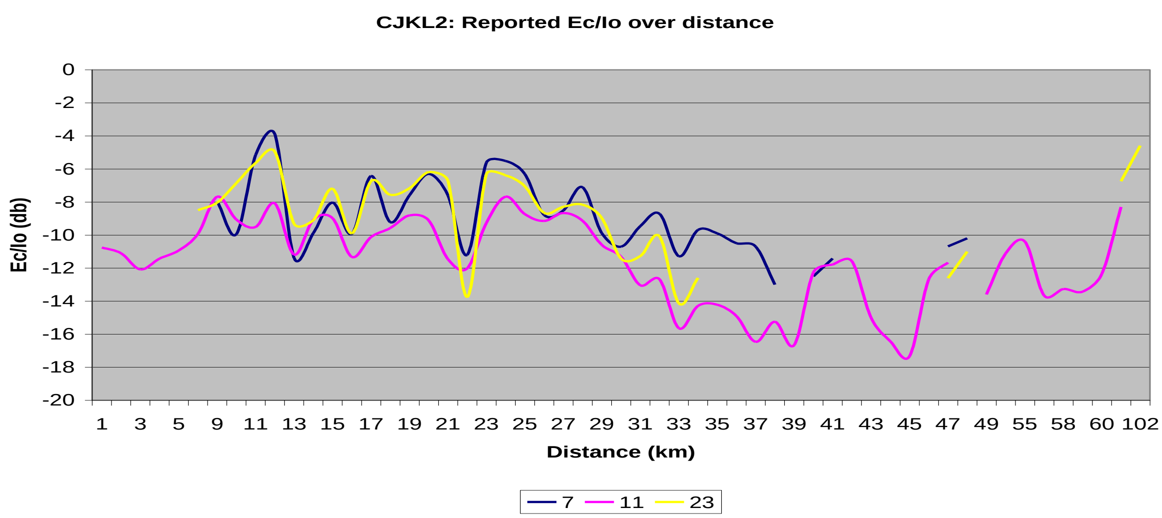


FIGURE 3-24.**Cell Breathing Impact.** This figure shows the Cell breathing impact on Ec/Io on a 24 hour period

It can be clearly shown that the reported Ec/Io during busy hour is weaker. In general there is about a 1-2dB degradation in the cell's primary coverage area (first 30km). The accuracy of the average Ec/Io improves if more call occurs (more calls results in more sampling)

I continue presenting my results by delineating the effects on the reverse and forward links, in this section. Table 3-10 presents the initial experiments performed in this study. The objective is to substantiate and understand the effects of cell-breathing, by means of the PCMD data, quantify the impact of cell breathing at the cell edge by analysing increase in noise, dropped calls and lost calls, etc.

Link	**Experiment**	**PCMD Data Used**
Reverse	Ec/Io observed vs. call count from the distance of the cell-center	Call Count Distance
	Calls dropped vs. distance from the cell center	Dropped calls Distance Hourly Measurement
	Calls placed vs. distance	Calls placed Distance
Forward	Ec/Io values vs call counts at various hours, at various distances from the cell site	Call count Distance Hourly measurement

TABLE 3-10. **Cell Breathing Experiments.** I separate the forward and reverse link in order to quantify the effects of cell-breathing correctly.

3.4.5 Reverse Link Measurements

In this section discussing reverse link measurements, I present the figures first and discuss the collective conclusions towards the end of the section.

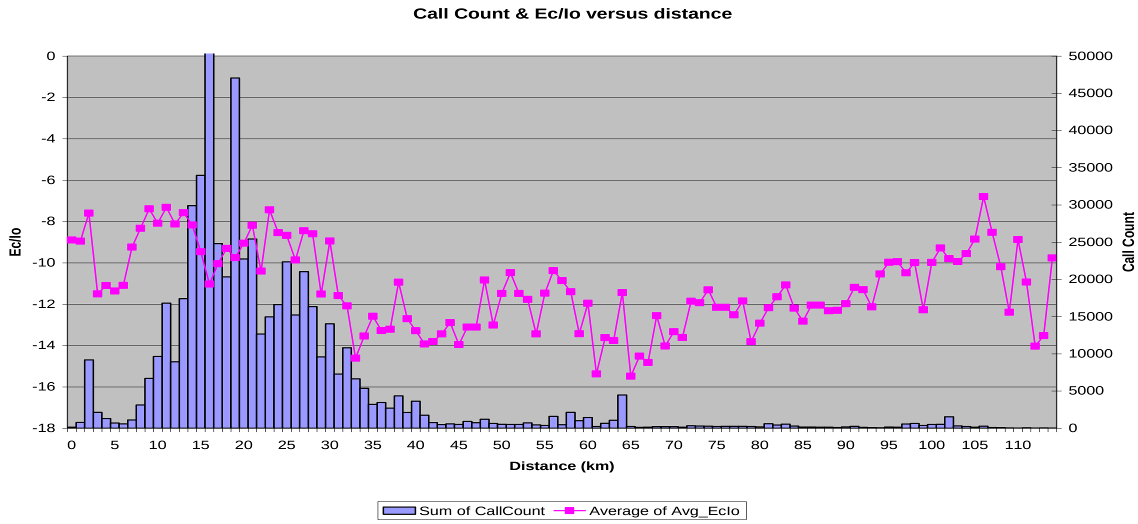


FIGURE 3-25. **Reverse Link Traffic**. Call count and Ec/Io vs. Distance from the cell-center.

Fig. 3-25 describes the Ec/Io values, shown by the solid line, on the reverse link, shown on the left hand side y-axis, against the call count (depicted by solid bars) as the distance from the center of the cell increases, along the x-axis. The distance from the center of the cell is measured in kilometers. The Ec/Io values drop significantly in the region where the number of calls is the largest, around the 15km-20km range, from the center of the cell. There is also a significant drop at the edge of the cell (90km-100km +)

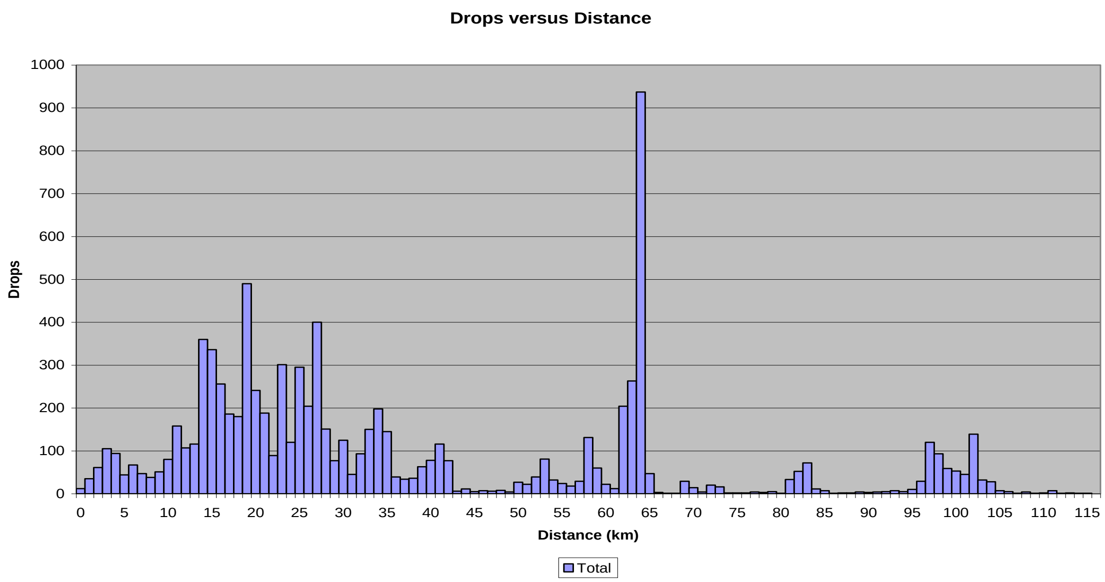


FIGURE 3-26. **Dropped Calls** . Calls dropped vs. Distance from the cell-center.

Fig. 3-26 describes another metric, number of calls dropped, versus the distance from the center of the cell. There are quite a few calls dropped at the 60-65km range and then the 17-23 km range, as seen in this graph. The y-axis shows the number of calls dropped and the x-axis shows the distance, in km, from the center of the cell. The effects are also prevalent at the edge of the cell (around 90km-100km from the center).

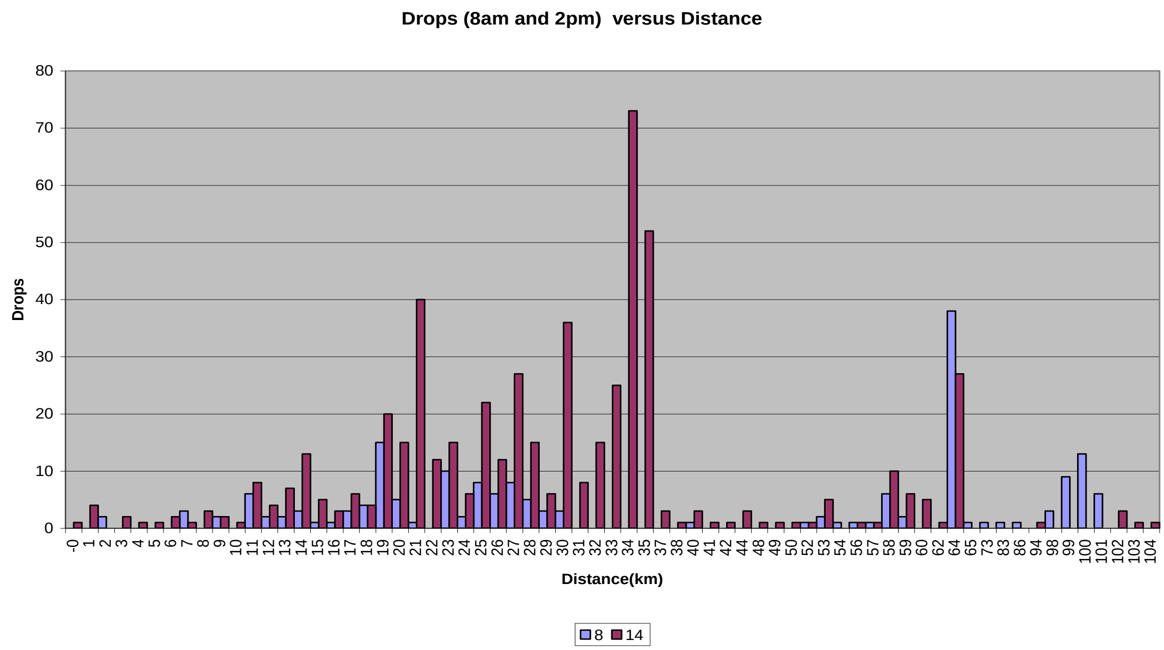


FIGURE 3-27. **Dropped Calls at 8am and 2pm** . Calls dropped vs. Distance from the cell-center.

Fig. 3-27 describes another metric, number of calls dropped, versus the distance from the center of the cell at two points in time, to contrast the differences. There are quite a few calls dropped at the 60-65km range and then the 17-23 km range, as seen in this graph. The y-axis shows the number of calls dropped and the x-axis shows the distance, in km, from the center of the cell. Two time points, 8am and 2pm, are plotted on this graph. The blue solid lines depict the calls dropped at 8am and the red solid lines depict the calls dropped at 2pm, to compare and contrast the load on the cell. The legend in the chart says 14, which stands for 14:00 hours or 2pm.

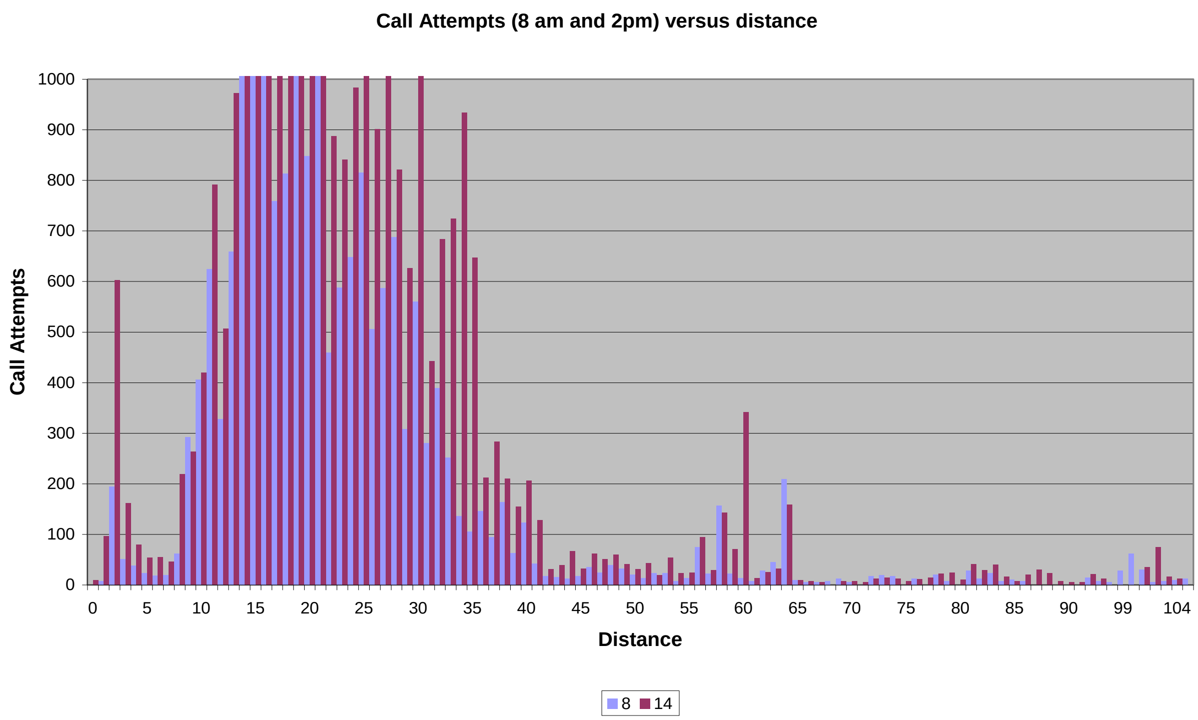


FIGURE 3-28. **Calls placed** . Calls placed at 8am and 2pm vs. Distance from the cell-center.

.

Fig. 3-28 describes the number of calls placed, versus the distance from the center of the cell. A majority of the calls are placed in the range of 10km-35km from the center of the cell, as seen in this graph. The y-axis shows the number of calls placed and the x-axis shows the distance, in km, from the center of the cell. The blue solid lines depict the calls dropped at 8am and the red solid lines depict the calls dropped at 2pm, to compare and contrast the load on the cell.

From Fig. 3-25 to 3-27 I observe that the cell is not Forward link limited (Ec/Io is strong even at >100km), as seen in Fig. 3-23. I expect users at the cell edge (e.g 90-100km) to experience a high number of dropped calls (in relation to call attempts, as seen in Fig. 3-24) . These drops are caused by the user being able to set up calls at cell edge but because the calls are extremely marginal (transmit power to max), the calls don't hold Ill and drops frequently (Fig. 3-25). Calls that don't get set-up at cell edge (obviously don't get recorded). Therefore most calls at cell edge would succeed when traffic is light (no cell breathing). (Fig. 3-26 shows several calls beyond 60km)

3.4.6 Forward Link Measurements

Fig. 3-29 shows the Ec/Io measurements at various distances from the center of the cell contrasted with the total number of calls placed at that distance. The calls at 7am and 12pm are compared in this graph to have two points of reference.

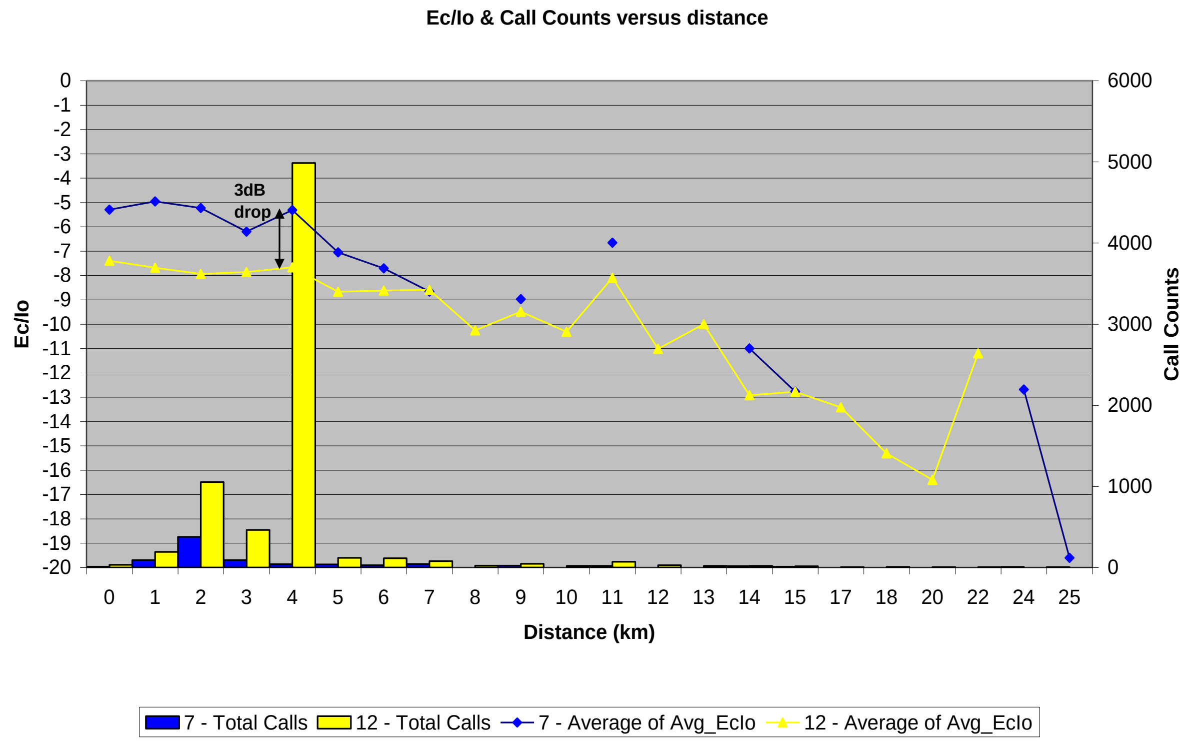


FIGURE 3-29. **Forward link signal strength.** Ec/Io and call counts vs. Distance from the cell-center.

The calls made at 7am are plotted in dark blue bars and the dark blue solid line depicts the Ec/Io at that time. The yellow bars and solid line is for the calls placed at 12pm. I observe that the deterioration of Ec/Io is uniform across the site (at various distances) The deterioration of Ec/Io is possibly 'uniform' across the entire cell at various distances of significant traffic. At distance of between 1km and 4 km (where calls are statistically sufficient), the deterioration is observed to be approx -3dB. This agrees with the theory that when traffic goes up, the interference is experienced equally among all mobile users making calls on that particular cell. Fig. 3-30 shows the Ec/Io measurements on the forward link, plotted using the solid line, vs. the total number of calls made (depicted by the solid blue bars) at every hour of the day. I see in this figure that the deterioration in the

Ec/Io is inversely proportional to traffic. As traffic increases (call attempts increase). The graph clearly shows that overall Ec/Io of the cell deteriorates as number of calls increases during the day (8->18) and improves after traffic reduction (>18). I observe from this graph that cell breathing occurs when significant number of calls are made at the site

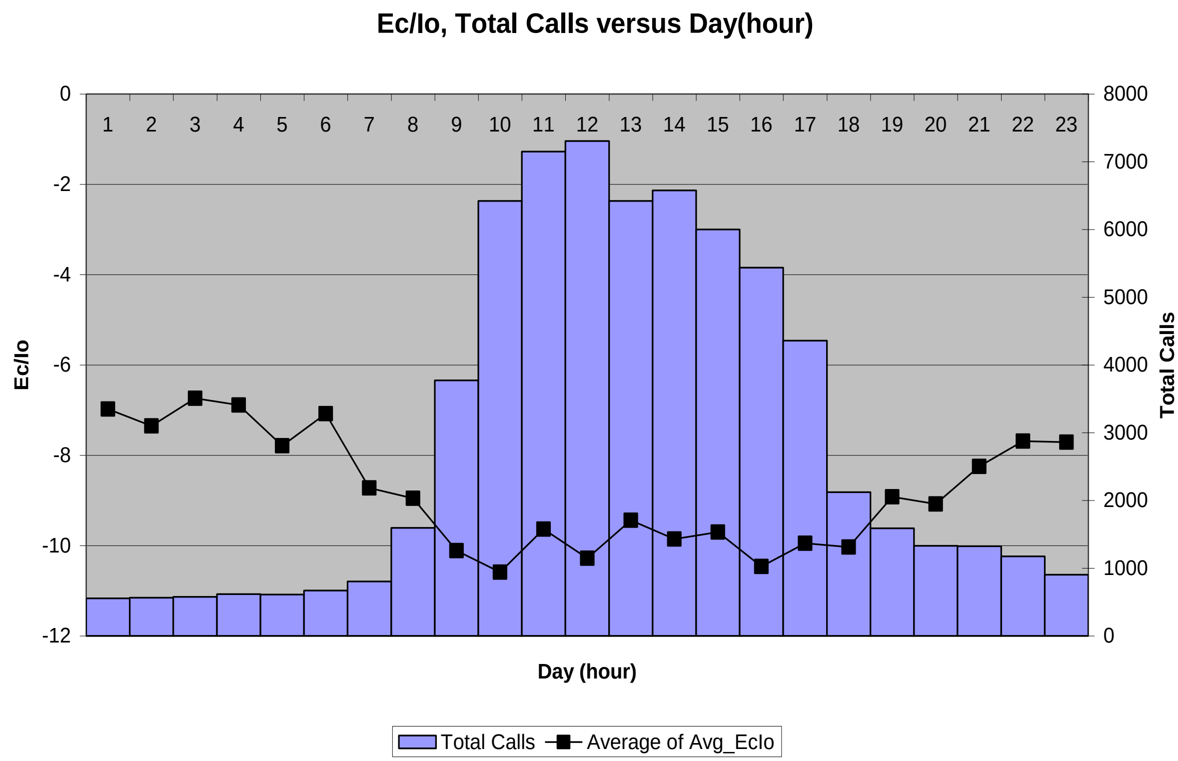


FIGURE 3-30. **Hourly Forward link signal strength.** Ec/Io and call counts at every hour of the day.

3.5 Related Work

Visualizing massive data-sets pertaining to networks, specifically telecommunication networks, is not a new problem [29]. Several end-uses of such visualization have been proposed, including city planning [30], analysing commute times and automotive diagnostics by means of mobile sensors [31], Internet services with voice network traffic [29,32], vehicular internet access [33], analysing social networks [34,35], real-time survivability metrics [36] etc. While some of these studies use data from sensors actually deployed on moving entities [31], many of these still use simulated data. Mathematical

Optimisation models have been proposed in order to investigate load-balancing problems in CDMA systems [37] for non-uniform traffic distribution. Genetic algorithms have been proposed to solve the problem of traffic hot spots and unbalanced call distributions [38]. Adjustment of pilot power has been proposed [38] in order to assist in better load balancing. A comparison of load-balancing metrics has been presented in [39] wherein a novel vertical handoff algorithm is proposed, in order to optimise performance across WLAN and CDMA networks.

The Vallent Prospect System consists of performance analysis tools for monitoring, fine tuning or troubleshooting an existing Flexent/AUTOPLEX Wireless Network [41]. It includes modules to enhance the operation, maintenance, and engineering of the wireless system. Modules can be used to monitor system performance, to report system configuration, and to diagnose, track, and solve many types of wireless engineering and system operation problems. A feature to trend SCME critical triggers is also available in the tool set. The feature provides pre-defined report templates to support trending critical triggers. The feature allows users to predict the trend of critical trigger data over time to see when a particular resource measured by a critical trigger will become exhausted. The trended intercept of the trigger with the limit is determined and displayed for individual network elements. Individual trend results are time-ordered by the projected intercept date and colour-highlighted in a summary report. [41]

3.6 Conclusions

In this section, I discuss the implications of my findings in section 3.2-3.4 on system design for performance engineers, who are responsible for day-to-day maintenance of QoS guarantees. Several tools exist [Prospect],[Periscope] that provide the engineer with visualization tools to enable better performance monitoring. These tools do not *combine data sources* across PCMD and service-level measurements, which could extend to such parameters as RSSI values. Combining data sources and providing the engineers with a coherent view of the network and subscribers motivates establishing user-mobility patterns, detecting how load-balancing works within the network and understanding cell-breathing. With the migration to UMTS networks [news reference], cross-network optimisation becomes an arduous task. Since the subscriber base remains the same, the patterns of movement provide insight into how such optimisations may be performed both from a provisioning perspective and the perspective of enabling new applications such as LBS and Emergency Management.

Service measurements targeted at detecting noise (RSSI over thermal noise) can be obtained from the appropriate measurement counters. This however is not based on PCMD data but based on service measurement data(hourly, by sector). Average RSSI Rise Above the Noise Floor is a count that provides the average Received Signal Strength Indicator, which is a measurement of power present in a received radio signal. In my data-set, the RSSI rise is generated every 2 seconds. The median RSSI is measured, sorted and calculated in the radio (CBR for modcell and BCR for SII, the radio measured the total receive signal strength in RX path.A filtered lowest RSSI is taken as the noise floor (from the RSSI measurement).The RSSI average is a filtered RSSI value based on the median RSSI reported by the radio.The average RSSI rise is averaged for an hour for SM (according to the averaged RSSI value taken every 2 seconds). The peak value is the peak RSSI rise (sampled every 2 second when blocking decision is made) for that hour. To determine whether there is interference monitor the Average RSSI (Received Signal Strength Indicator) Rise Above the Noise Floor, a good indication of interference is if the count goes above 300. (divide by 100 to get dB), which is 3dB. Anecdotally, if the average rise is consistently higher than 300, external in-band interference has been discovered by many customer markets. Note this may be due to inter-modulation products (mixing of signals from different transmission sources and/or faulty components in the transmit or receive path within the cell), or in-band transmissions from other RF transmitter sources. In a study of the busiest site in North America, one close to the Giants Stadium in New Jersey, it was observed that the RSSI averages were off the charts. In the duration of the game, this average was even worse. In my study, I choose to compare and contrast the load-balancing patterns that are seen on the dates of an All-Blacks rugby game using a different combination of parameters, Ec/Io, caller traffic and caller *postion.* The results show that load shedding is biased in the direction of user-movement and therefore, this is a valid variable to analyze. I make note of the point that CDMA presents inherent advantages to other technologies (GSM, TDMA) in that it is able to accommodate dynamic traffic patterns. In the other technologies, blocking might occur based on power and RSSI limits being reached with an increase in caller traffic (and this is exemplified by the dates on which major tournaments, matches and games occur and can be extended to situations such as emergencies where the first thing people do is call someone). Robust RF design also contributes to managing call traffic Ill. Since the users are mostly static when the game is being played, even with an increase in the average RSSI due to increased caller volume, an individual callers path-loss is not at its maximum value. Another factor that is critical in good load balancing is the base-station itself, which should support inter-carrier load-balancing. One of the biggest concerns in emergency situations is handling the case where a cell-tower goes down. It would be beneficial to understand the effects at the center and at the periphery of the cell, as established by my cell-breathing studies, in order to establish patterns relating to how call performance is affected or needs to be re-assigned amongst adjoining cells.

4 Reputation

In order for *proactive crowd sourcing* to work well, one element beyond location that deserves close consideration is the notion of *reputation*. When asking citizens on the ground to assist with emergency response and management, it would be important to establish whether the information is coming from a reputed source. In this section, I expand on the ideas around reputation systems and present my benchmark for Social Networks, TESS. One of the ways in which a person's reputation can be established is by looking at their friend list, giving rise to the idea that the emergency manager knows *where* the citizen is, that is providing critical information and also *who they know* and becomes essential with the escalating severity and scale of the emergency. Section 4.1 discusses reputation systems today, section 4.2 describes the need to benchmark social networks, section 4.3 shows how confidence is assigned based on four metrics, time, expertise, sample size and source of the data, section 4.4 discusses the results of the assignment, section 4.5 details some related work in this area and section 4.6 discusses the conclusions and future work.

4.1 Reputation Systems

The simplest example of where a person's reputation comes in handy is in the financial sector, for loans, where a *credit rating* often decides a member's privileges and rates. The notion of reputation has been extended to various systems, as the nature of distributed computing requires assigning a basic level of trust to the source of the information and the information itself. Just as credit-ratings help a bank decide on how to treat a customer's account, reputation scores are established in order to assist in decision making, when it comes to using the services of an entity. Reputation systems are often collaborative in nature, where the comments of the crowd add or take away from an entities reputation. Reputation scores are a dynamic variable and are a function of time and incoming ratings. These systems find examples in commercially successful services such as recommender systems [97, 98] and collaborative filtering [99]. Assigning trust extends to other areas of technology including Peer-to-Peer systems [100, 101], establishing trust in online social communities [102, 103], online auctions [104] and *working* in online markets [105]. These ideas have found use in systems that are in use in everyday life, as shown in Table 4-1.

Technology	**Product**	**Metrics and methods**	**Assignment of trust**
Search	Page Rank	Number of forward links	They assign a page-rank of zero to sites using dubious advertising techniques and these sites are considered to be *spamming* the web-space.
	Blog search engines	Number of occurrences of author's name, citation analysis	Full-string matches.
e-commerce	eBay	Feedback chain, tag-clouds, social text mining	Feedback auctions are popular amongst deviants, amounts to what is essentially a Sybil attack.
	Epinions	Member opinions on products	Members can choose to "trust" or "distrust" each other, making this a more closely monitored social network.
	Bizrate	Agent based trust models	**Not clear (**A demonstration of the Agent Reputation and Trust (ART): testbed for experimentation and competition)
Identity Management	IDOlogy and Trufina	Matching information against public data-bases from the government	Assumes that the top level is the oracle
Social News	Slashdot	Bottom-up filtering of contributors	
	Digg	Votes from a community	Filter out multiple votes from the same IP or from proxies
Wikis	Wikipedia		Experts with established credentials are asked to review material

TABLE 4-1.**Commercial examples of reputation systems** The notion of reputation is used in several online systems and communities wherein each system uses a different metric or method to assign reputation and trust.

One of the most common attacks on reputation systems is the *Sybil Attack* [106, 107, 108, 109] wherein multiple pseudonyms are created for the same entity in order to tip the vote in favor of the malicious module. One solution for this has been proposed with SybilGuard [110] wherein the protocol borrows ideas from social networking, such that user-identities are connected by what is essentially a social network wherein each edge represents a human-validated connection or what they call a *trust relationship.* The model to defeat malicious nodes is therefore to place emphasis on trust relationships rather than the multiple Sybil identities that the malicious node might create. This gives rise to my argument that in the case of enabling emergency response by contacting mobile phone users co-located with the emergency, it would be essential to understand not just where the user is but also who the user knows, thereby adding a degree of trust to the transaction. On limitation with social networks however is that there is no reasonable benchmark to identify the validity of the results generated by the studies on these social networks. In the following sections, I describe a benchmark for Social Networks, TESS, which seeks to add confidence to the outcome of Social Network research, which is an integral component in establishing trust and *reputation* to the actors in my framework that provide information about an emergency.

4.2 Benchmarking Social Networks

Social Network analysis is used to understand the social structure, which exists amongst entities in an organization. The size, diversity and ubiquity of social networks act in combination to necessitate understanding these networks in a systematic manner. Several aspects of social network analysis are currently the subject of academic research. Some of the themes of research in social networks includes processes in on-line social networks relating to communication [93], the formation of communities [78], visualizing social network data [56], extracting social network metrics [76] and enabling various functionalities within social networks [84]. The data-sets used in these analyses are important to understand for the following reasons:

a. Data-sets used in any line of research ultimately lead to the formation of benchmarks, which are used to evaluate new proposals to address known bottlenecks.
b. Data-sets have to be accurate and representative of the problem being addressed, in order to provide confidence in the research being conducted.
c. Ultimately, understanding the nature of data-sets is required to perform accurate social-network simulation.

Social Network research has a multitude of experts participating from fields as diverse as Sociology, Anthropology, Computer Sciences, Library Sciences, Engineering and Information Technology. Bringing together such diverse expertise is not without challenges, especially when trying to understand which data-sets can provide results with the most confidence. In my initial survey of Social Network literature, I found a variety data-sets used to validate research, ranging from 14-25 user interviews to a 9-month survey of the users of Friendster ranging in the order of several thousands [56]. Depending on the question being answered, which varies across several categories, the nature of the data is bound to vary. This thesis proposes a basic outline of the characteristics, which the data-sets might need in order to make room for discussion across the board on the various research topics.

At first, I propose that four simple characteristics be taken into consideration. These characteristics include:

Temporal nature of the data: Social networks are experiencing growth similar to that of the internet. Over time, the growth experienced by social network in terms of raw size gives rise to new issues and perspectives, when it comes to the proposed solutions. Understanding the nature of this growth and having the data-sets reflect the temporal component of such growth is therefore imperative in this line of study.

Expertise of the participants: In the set of research papers, which used participant surveys or input, the expertise of the participants varied from being random participants [57] to extremely focused work-groups of GPs [58], AI researchers [59], etc. Once again, depending on the research question being answered by the research, I propose that the expertise of the participants is a factor, which will impact the confidence of the results produced.

Sample size of the data: When studying metrics relating to social networks, in particular, the sample size of the data-sets is an important factor. This extends to other analyses in privacy and trust, collaboration etc. as the size of the typical Social Network is always increasing through the addition of a global demographic of users, who wish to stay connected.

Source of the data: This refers to the background within which the data collected initially resided. The diversity of data-sets spans portions of popular Social Networking applications such as FlickR, Yahoo! 360 and Friendster [60] to a collection of conference papers [61] to e-mail lists [62] to wikis [63] to simply users carrying a certain type of cell-phone [64], understanding the source of the data is important in assigning a confidence metric.

The sum-total of this proposal is the evolution of a framework that incorporates the desirable characteristics T(emporal), E(xpertise), S(ample-sizes), (S)ource (**TESS**), which summarizes how well the data-sets used in a particular study relating to Social Networks. By analysing the data-sets used from the focal point of these characteristics, I go on to assign a confidence metric with TESS. An additional characteristic, which I hypothesize as being important, in certain areas of research, with varying definitions of metrics, such as centrality, trust etc. is the definition itself.

4.3 Data-Sets in Social Network Research

In this section, I discuss the methods I use to define and assess confidence in the data-sets used in Social Network research. The major themes I have encountered include extracting social network metrics, community formation, visualization, trust and privacy. I go on to present the metrics, which are a part of TESS, ultimately used to assign confidence to the answers proposed with the use of these data-sets.

A number of themes exist within Social Network research. Examples of these themes include, extracting social network metrics, community formation, visualization, understanding trust and privacy, collaboration, wearable computing and value-added services such as tagging, for online Social Network applications. In this section, I explore the broad categories of academic work within Social Network research and provide the characteristics of the data-sets used. This will lay the foundation for assigning ratings using the TESS framework in later sections.

4.3.1 Extracting Social Network Metrics

Social network metrics such as degree, between-ness, closeness and network centrality are often the subject of academic research. Understanding social networks and their metrics is important as these networks form the underlying structure, which allows for rapid information distribution [76]. A preliminary analysis of research includes data-sets from a variety of resources including email lists [62], the world-wide-wed [62] and Instant Messaging Populations [77]. Further, Social Network Mining using Google and data-sets from conferences [59] have been proposed to extract relations between people and identify groups. Table 4-3 presents sample data-sets, which are used in these studies. [78] suggests the use of computer-generated networks, to perform a controlled study of metric extraction and the use of bibliographies from arxiv.com to study this problem. Zachary's karate club network [120] is used in this work to understand the real-world applications of these ideas. Link topologies have been used [134] to predict social connections and extract metrics indicating connectedness. Information sharing has been proposed with the use of Saori [80], in order to enable information dissemination. A relationship algebra [81] has been proposed, in order to understand and analyze social connections using data-sets from publication bibliographies and parts of the online network Orkut. The Citeseer dataset is used in [82] to understand how the social actors, in this case authors of various papers, affect the lines of research, which are observed. A new research paper search engine, Rexa.info, is proposed in [83], in order to organize publications for effective retrieval, enabling social network analysis. Event and place semantics are extracted using Flickr tags in [84] to extract usage patterns of people sharing photographs. Table 4-3 summarizes the data-sets used in these studies.

Year of Publication	**Data Set**
2005	4 academic conferences, 500 participants, 3 years
2004	53 e-mail participants, 229 web-pages
2004	Buddy lists from Live Journal, 25 days
2005	1 academic conference, 503 attendees,
2000	145 scientists, bibliography over 3 years
2000	1265 people, Friends listed on personal homepages in Stanford and MIT
2006-2007	49897 photos from Flickr.com, 1015 days worth data
2000 -2001	108,676 academic papers from Citeseer, 13 years worth of data

TABLE 4-3.**Sample data-sets from Social Networking research** I present some data-sets used in Social networking research, classified by the year in which the research was published.

4.3.2 Community Formation

Community formation is important to understand within Social Network analysis, in order to understand patterns of collaboration. BitTorrent communities were studied in [85], in order to understand the factors affecting the paticipant's co-operative behavior. The Iris and DPLB datasets are used in [86] to mine communities within social networks. Group Formation is studied using data-sets from LiveJournal and DBLP [87] to understand the evolution of communities. User experiences at Open Office [88] were discussed wherein an open source office suite with nearly 62,000 mailing list subscribers was analyzed. Data from a hundred mobile phones were analyzed in [64], over a period of nine months, in order to understand and reflect on social patterns. A user-group of older people [89] was used to understand the accessibility and inclusion of this demographic, in online social interaction. An online community in a suburban town was studied [90] to investigate means to stimulate social engagement. New information interfaces are proposed in [91] to provide hypermedia capabilities for information sharing and collaboration. Blog entries are mined [92] to discover stories within the data found in blogosphere. Digital Libraries act in unison [93] to create a common learning substrate accessible by a variety of learners with a proper interface to stimulate learning. [94] uses two sets of data, both synthetic and real-world data, to identify communities, while proposing heuristics to analyze what could be NP-Hard problems. Table 4-4 summarizes the data-sets, which are used in these studies.

Year of Publication	**Data Set**
2006	875 LiveJournal communities over 10 days, 71,618
2005	70 Conferences
2006	62,000 registered
2005	100 people using Nokia6600
2002	280 individual visitors
1997	colleges, 15 teachers and administrators,
2006	1200 on-line messages
2007	Two synthetic data-sets (Assembly Line, Dutiful Children),two real-world data-sets (Southern Women)

TABLE 4-4.**Community Formation Data** I present some data-sets used in the sub-field of community formation research across various years

4.3.3 Visualization

Visualizing social networks assists researchers in understanding new ways to present and manage data and effectively convert that data into meaningful information [95]. A number of tools have been proposed for this task of visualizing social networks including Pajek [96], NetVis , Krackplot, IKnow, InFlow, Visone, JUNG and Prefuse, to name a few. Discussion forums are considered to be another source of online collaboration and these have been visualized to better understand interactions [97]. Visualizing tasks for better collaboration during software development has been proposed [97] to address issues of co-ordination and geographical distribution of developer teams. Visualizing social networks using Query interfaces for wikis and blogs [63] are used to provide the end-users with more user-friendly alternatives. Weblink graphs were used in [98] to extract hierarchies of complex networks. Over a year-long period, individual and team use of tablet PCs was studied [99] to understand the process of learning, within a group of students. Business Intelligence search [100] was facilitated by looking at the agreement between participants on certain statement and visualizing the same. The network value of customers is visualized [101] in order to enable direct marketing more effectively. Visualization further finds its use in law-enforcement [102], wherein crime-pattern recognition and criminal associations were mined and visualized for the Tuscon PD. Bibliographic data has been visualized, which finds its end application in summarizing scientific fields. Simple techniques for visualizing social graphs [103] have been proposed. Table 4-5 summarizes the data-sets, which are used in these studies.

Year of Publication	Data Set
2007	Discussion forums, 16 participants
2006	Web link graph of 51,497 internet pages, An empirical set of 8210 word-associations
2006	7 participants using tablet PCs over a 12 month period
2007	30 undergraduate researchers assessing two websites to gain business intelligence
2005	Incident reports and GIS tools from the Tuscon Police Department
2007	2.8 million movie ratings of 1628 movies by 72916 users over 18 months
2002	Dataset from the 2001 Graph Drawing Contest with papers from 1994-2000
2004	5 students creating messages for one another

TABLE 4-5.**Visualization** I present some data-sets used in the sub-field of social network visualization across various years

4.3.4 Trust and Privacy

As the size and ubiquity of social networks grow, trust and privacy become very important issues for both designers and users to address and understand. The wordpress blogging engine is used in conjunction with Mozilla Firefox, in order to provide signature-based architectures for secure communication on the Social Web [104]. The MovieLens data set is used in [105] to deconstruct recommender systems. Extensions to the RDF framework to incorporate mechanisms to enable trust have also been proposed [106]. Other examples of data-sets include MBA students from colleges in the USA [107], supply chain data [108] and agents [109]. Over 1200 people from EU countries were studied [110] in order to analyze the value of location privacy. Community connectedness as understood by analysing the privacy requirements [111] was studied in order to reduce detachment within online social communities. Sybil attacks in distributed systems wherein several fake identities are utilized to start attacks are studied [112] and countered using trust, which is introduced by adhering to social networks amongst user identities. The value of creating on-line identities has been explored [113] in order to understand identity theft in online communities, which actively encourage users to create profiles, share personal information and network socially. Table 4-6 presents data-sets used in this line of research.

Year of Publication	Data Set
2005	18 to 706 profiles of movie ratings from MovieLens
2000	239 students evaluating online shopping system
2002	53 e-mail participants, 229 web-pages
2002	Supply Chain vendors
2006	2000 participants from law, sciences, ten European countries carrying mobile phones, 1 year
2007	18 users of environment services, 80 percent expert users
2006	A synthetic social network model with 10,000 nodes
2006	41 respondents in a classroom setting

TABLE 4-6.**Trust and Privacy** I present some data-sets used in the sub-field of privacy and trust across various years

4.4 Comparison to the Southern Women's data-set

The Southern Women's data [75], albeit being a relatively small and old dataset, encompasses some of the necessary characteristics in data-sets, which can be used to confidently assess new proposals within Social Network research. These characteristics include the following:

4.4.1 Temporal Characteristics

The Southern Women's data was collected over a 9-month period. Since the size and popularity of Social Networks is growing in leaps and bounds, in order to address the problems or roadblocks in the development of these networks, it would be essential to have data-sets which are characterizing the networks over a period of time. A single snapshot in time may not be the most effective way to gather data, to assess new proposals for research problems.

4.4.2 Expertise

Sampling the data from various points of expertise is important because this normalizes the confidence in the data. For example, some datasets are procured from researchers with several publications [59] while some other datasets are procured from members of a fraternity house [114]. These ranges of data need a meeting point by assigning a certain degree to the expertise characteristic, in order to enable fair evaluation. In the Southern Women's data, the expertise of participants also varied between three peer-groups, the women who participated in the survey, observers and members of the press.

4.4.3 Source

The source of data is important to understand with the rapid proliferation of information. All data sources from Wikis to proper bibliographies are represented in the data-sets popularly used in

Social Network Research. There needs to be a clear understanding of the source of the data, in order to understand whether the data is indeed reliable.

4.4.4 Sample Size

The Southern Women's data set had only 14 participants. The data-sets analyzed in this thesis vary between a few participants to several thousand participants. In order to be confident in the results of any proposal, the sample size of the data needs to be analyzed and discussed, in order to realize its suitability for the analysis.

The TESS framework is simply aiming to assign a confidence metric, based on the data-sets used in the research. I am using a simple assignment of ratings between 1-5, 1 being the lowest and 5 being the highest confidence. This creates a confidence-vector, which is TESS. It is desirable to have a confidence-vector, which balances all the four characteristics equally or pinpoints exactly why a certain rating for any of the characteristics is where it is.

4.5 Assigning Confidence Vectors

Using the metrics outlined in table 4-7, I assign a confidence factor to the literature pertaining to Scoial Networks, where a diverse set of data is employed to address several existing challenges. Fig. 4-2-4-4 shows the confidence-vector, TESS, for the fields of social network extraction, community formation research and trust and privacy research. I choose to leave out visualization as the basis of this study would be better performed if juxtaposed with the multiple actual tools [96], which exist, several of them being open source. The x-axis pertains to the literature whose data-sets I analyzed and the y-axis shows the actual confidence rating, in each of the characteristic fields within TESS.

TESS Confidence Ratings for Data-sets	
Metric	Confidence Rating
Temporal Characteristics	
Single snapshot in time	1
Measured over several hours	2

Measured over several weeks	3
Measured over several months	4
Measured over several years	5
Expertise	
Random participants	1
Anticipated users of portion of the technology	2
Actual users of the technology	3
Authors or developers of technology	4
Domain experts, for ex. data-sets taken from conferences	5
Source	
Single source	1
1-2 Sources	2
2-3 Sources	3
3-4 Sources	4
4 or more Sources	5
Sample Size	
Less than 100	1
100 - 1000	2
1000 - 10,000	3
10,000 - 1 Million	4
Greater than a Million	5

TABLE 4-7.**Confidence Rating.** I present how I rate the various data sets based on their characteristics such as being measured over a small/large period of time, the expertise they present, the number of sources they look into and the sample size that they consider.

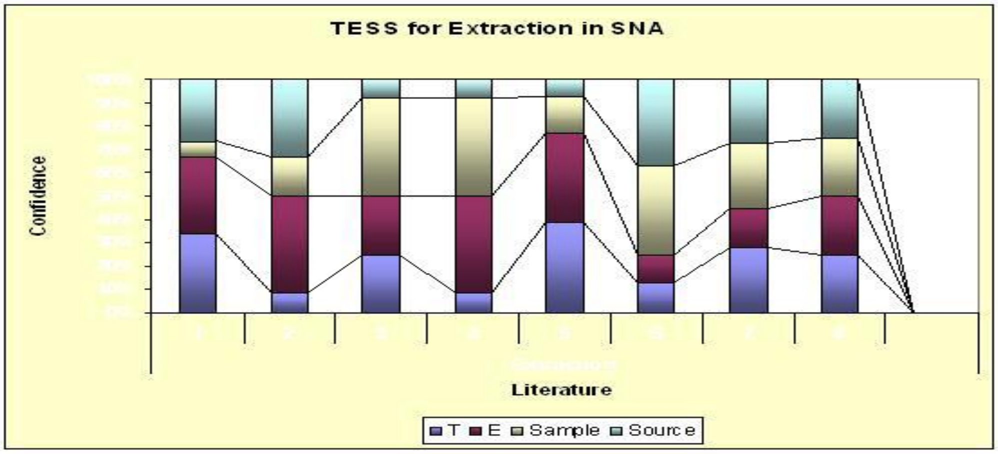


FIGURE 4-2.**Social Network Extraction.** TESS ratings for Social Network Extraction

In Fig. 4-2, the average rating varies between 3-4 in this field of Social Network analysis, that of metric extraction. The expertise of the participants is rated the highest in this Field of work as the data-sets mostly pertain to academic conferences.

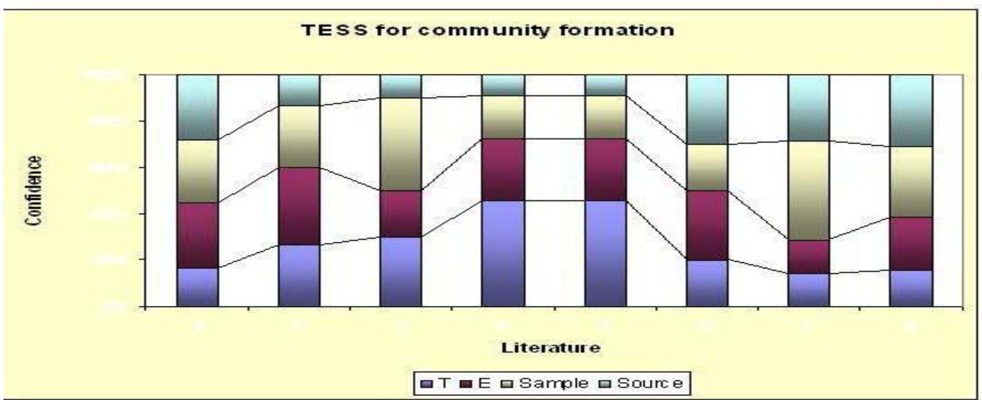


FIGURE 4-3.**Community Formation.** TESS ratings for the formation of communities

In Fig. 4-3, the average rating varies between 2.375 to 3.35, with the source being rated the lowest. This is because the source of the data-sets in this field of study seem to pertain to a single demographic of users, such as users of a certain model of phones or mailing-list subscribers.

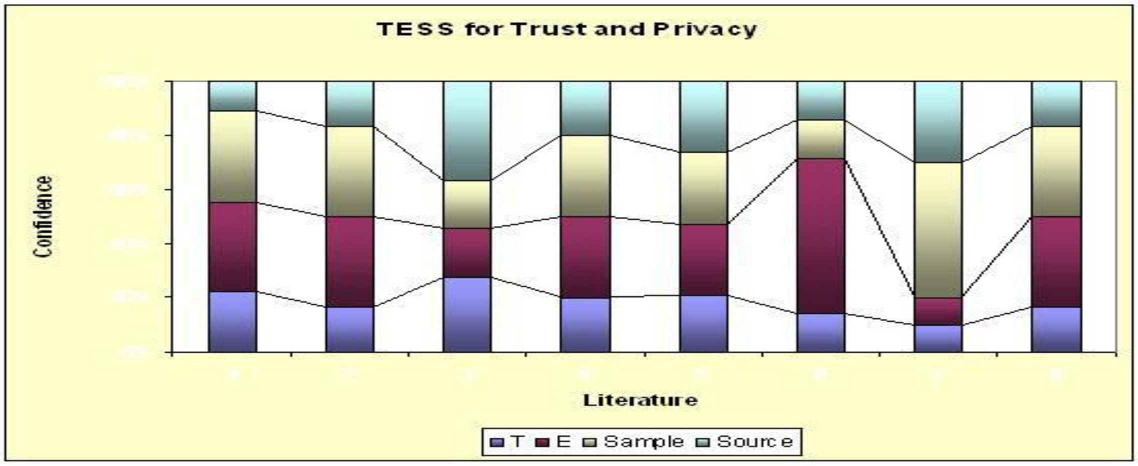


FIGURE 4-4.**Privacy.** TESS ratings for Trust and Privacy within Social Networks

There are no peers from different demographics to even the ratings out. In the Southern Women's data, for example, besides the participants themselves, the sources included the press and observers, representing different peer groups. Within studies on community formation, it seems that the analyses could extend to providing analogies of the utility of the proposed solution within different demographics.

In Fig. 4-4, I see that the ratings are the poorest across all characteristics, ranging between 1.8-2.8. The temporal aspect of the data-set is valued at the lowest in this set of data because most studies seem to consider a single snapshot in time. This observation could also be attributed to the fact that research pertaining to trust and privacy, mostly propose alternate models [108] for trust or improved security in terms of protocols or alternate specification [106]. In this case, implementing the proposals on data-sets spanning time may or may not be relevant for initial analyses. This leads us to a discussion of factoring in the definitions of various metrics, into my assignment of confidence, which is the subject of future work. This implies that the actual definition of trust, privacy and other metrics such as centrality, might affect the assignment of a confidence metric to data-sets.

4.6 Related Work

Network data, in particular Social Network data is available from many different sources. For example, some of the data-sets used in network analysis include; Zachary's karate club [65] defines a social network of friendships between 34 member of a karate club at a US university. The co-appearance network of characters in the novel Les Miserables [66] has been created and the adjacency network of common adjectives and nouns in the book David Copperfield has also been studied [67]. A network of American football games between Division 1A colleges during the Fall 2000 season [68], an undirected social network of 62 dolphins in a community from Doubtful Sound New Zealand [69], a directed network of hyperlinks between blogs on politics in the United States recorded in the year 2004 [70], a network of books about US politics sold online by Amazon.com [71], and a network of co-authorships of scientist working on network theory and experiment [72].

Several benchmarking schemes exist in the area of Knowledge Based Systems, which can be extrapolated to the Semantic Web and further to Social Networks. The Resource Description Framework (RDF) is a family of W3C specifications, which has become an accepted form of metadata, extending itself to Semantic Web applications in such manifestations as RSS and FOAF ontologies. The FOAF ontology is machine-readable and used to describe people and their interrelations with other entities. This ontology decentralizes the data used in Social Networks by allowing users to create and describe social networks, without referring back to a central database. FOAF extends the RDF specification and is described using OWL. The Leigh University Benchmark (LUBM) is used to benchmark the Semantic Web with respect to use in large OWL applications. The LUBM uses a uniform ontology and can be applied to various scales and configurations. Benchmarks further exist for specifications such as Web2.0, used in extending the social semantic web. For example, del.icio.us can be considered as one such benchmark for Web2.0.

Since Social Network research embodies a range of expertise from anthropology to Computer Sciences , it is difficult to find benchmarks for social networks, per se. This thesis aims at analysing the data-sets used in various fields of Social Network research to perform the groundwork for such benchmarking in the future. Social Networks have been measured in many ways and the measurements have been carried out on various data-sets, from on-line social networks [73] to sexual transmission networks [74].

Some benchmarks are known in social-networking literature including the Southern Women data from 1985 [75]. This particular data-set dates back to the 1930s and was used to understand inter-personal relationships. Using sociological definitions, researchers such as

Roetlisberger and Dickson (1939) and Davis, Gardner and Gardner (1941) segregate the data into core and peripheral group members. In the 1930s, five ethnographers collected data pertaining to stratification, from Natchez, Mississippi. The aim of this study was to understand social class in a mixed-race society. Eighteen women were picked for this study and a systematic analysis was carried out on their social activities, over a nine-month period. During that time, subsets of this set of women participated in social events and their participation was monitored by means of interviews, recording the observation of other participant observers, guest lists and newspaper reports. The size of this data-set was small and the source was not definitive, given two contrasting studies. The size of social networks has grown considerably over the last seven odd decades, making the size of the sample set more relevant. The temporal aspect of the Southern Women data-set is to be noted. The researchers did not simply sample one point in time, in order to carry out their analysis, but in fact sampled the data over a nine-month period. The expertise of the data-set here was not limited to the participants themselves but included that of the observers and the press. This data-set, while from the 1930s, certainly took into account the inclusion of the basic characteristics, which I propose as part of this work, in the TESS framework.

4.7 Discussion - Assigning reputation to mobile phone users

In this section, I develop the idea of assigning a reputation value to mobile phone users. A number of variables could be taken into consideration but, I focus my discussion to the metrics that could be applied to the commercial applications such as LBS and the non-commercial applications such as Emergency Management. I use the PCMD data-set discussed in detail in section 3 and an associated data-set, the *subscriber handset data-set*, which identifies which handset a particular subscriber is carrying, based on the Mobile Identification Number (MIN).

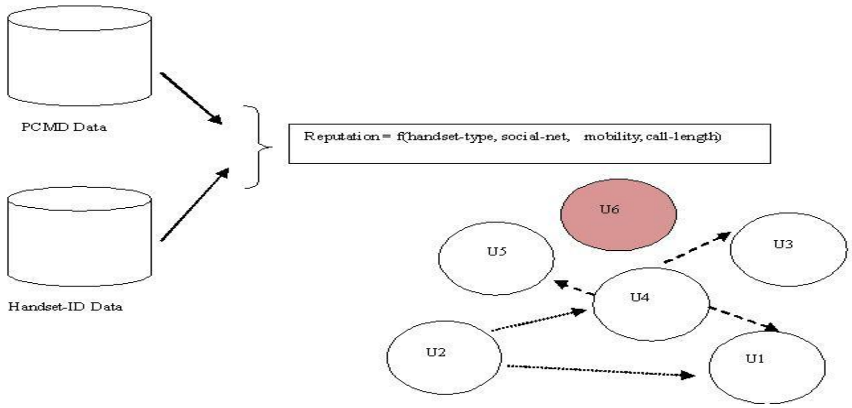


FIGURE 4-1. **Assigning trust to mobile-phone users.** Using a combination of PCMD data and data pertaining to handset identifiers, I assign reputation based on four variables, handset-type, social network, mobility and call length.

I propose a reputation metric that is based on four tuples, handset-type, social network of the caller, mobility of the caller and the average length of calls that a user makes. I explain these variables below, in detail:

Handset-type: As outlined in section 2, the handset a user carries determines the price-point and the features available on the device such as a high-resolution camera etc. With the proliferation of cheap SIM cards, one of the easiest hacks by a malicious party at the time of an emergency, to avoid being tracked would be to simply change the SIM card. A Mobile identification number is a more reliable source of information about the user.

Social Network: I use a couple of fields in the PCMD dataset in conjunction to establish the social network of the phone users. These are listed in table 4-2. The social network of a caller can be established by looking at fields 22 and 52 and constructing a graph. A geographical network of users can also be created by analysing the initial and last cell sites where the calls are placed.

Field number	**Length**	**Verbose output field name**	**Field description**
1.	16 bits	PCMD Version	Identifies data layout. Unique for each release.
3.	32 bits	Call start Time	Time call was started Set according to the access or page response message as follows:
4.	32 bits	Call Length	Call elapsed time The RF time of the call. This is calculated from the new call start time and the time of normal or
5.	6 bits	Call type	Call processing call type
7.	20 bytes	Subscriber number	Subscriber directory number

8.	10 bytes	MIN	Mobile identification number
10.	16 bits	Ini. Cell Site Num	Initial cell site number
14.	16 bits	Last Cell Site Num	Last cell site number
22.	20 bytes	Originated Digits	Number the person is calling
52			Number calling the person (incoming)

TABLE 4-2. **PCMD fields used to assign reputation** I create the social network of the user by looking at the numbers calling them and the numbers that they call. Further, the PCMD data also records other parameters of relevance that can be used to inform my reputation metrics. I create the social network of a unique subscriber by profiling the numbers that they call and the numbers calling them.

Average length of calls placed by the user could potentially indicate, in the case of an emergency, the purpose of the call (a quick call to see if their friends are OK or a longer call to establish contingency plans, etc.).

Distance Travelled by the user might indicate how far from the site of an emergency and in which direction a user is travelling, in order to establish how much useful information they can provide.

4.8 Conclusions

I envision that TESS will be used by social network researchers, when presenting the validity of their results with varying levels of confidence. Depending on the nature of the problem being addressed (or the sub-problem such as community formation, etc.), the ideal TESS vector can vary. Additionally, TESS lays the basis for the metrics that should be considered while assigning *reputation* including the temporal nature of the data (over what time period is the data collected), the expertise of the participants (perhaps this co-relates to how many services a cell phone user actively uses or what sort of handset they carry), sample-size (how deep is the individual's social network, how many individual callers are we contacting the case of an emergency) and source of the data (are we contacting ordinary civilians alongside special emergency responders?).

5 Reverse-111

In this chapter, I demonstrate the use of Reverse-111, a *proactive crowd sourcing system* which takes into account the fact that several users might be co-located with an emergency and can be contacted by an Emergency Centre, to provide relevant and useful information pertaining to the situation. I have co-developed a variety of first responder agents who can participate within a response simulation along with the PARVAC team at the University of Washington.

The Reverse 111 tool extends the RimSim Response architecture [117] that has been developed as a visual analytics project, which enables first responders to plan and train for potential emergencies and crises that might arise in their area of responsibility. This architecture was built primarily to enable the notion of synchrony amongst the various administrative units, in responding to an emergency. The Emergency Response Centre personnel are required to manage situations on a day-to-day basis and participating in a variety of drills prepares them for scenarios that are unforeseen. By bringing in an element of asynchronous training opportunities, RimSim Response offers computer-based participation exercises for the first responders. This gives the participants several tools to understand each other’s roles, actions and thought processes. Complex decision-making scenarios are implemented in these simulations in order to simulate the scale of disasters in a place like Seattle. Various situations can be simulated including bio-weapon attacks, tsunamis and other classes of attacks. Several bodies including the National Incident Management System (NIMS) [118] provides first responders with guidelines for activities to train personnel in. A lack of real-time data and an increase in complexity when it comes to distributed management provides a sizeable challenge in this area. Role-playing and table-top games have been used in several placed but RimSim Reponse overcomes the shortcomings of these training tools by means of online simulation to understand distributed cognition.

I propose the use of four different agents, greedy, round-robin, local and lottery that implement response heuristics and communications strategies to enable a successful community response to crisis. I develop a tool for New Zealand, which simulates disasters at various intervals and various geographic locations, that are potentially tied to different administrative domains and I present my initial findings from interacting with multiple agent simulation sessions. Section 5.1 introduces my proposal, Section 5.2 describes my agents in detail, Section 5.3 presents the visualization interface that I developed to allow the agents to allocate resources to incidents, Section 5.4 presents the performance results of the different agents, Section 5.5 talks about related work and Section 5.6 presents my conclusions.

5.1 Reverse 111 – Proactive Crowd Sourcing

Reverse 111 proposes *co-locating users with emergencies and obtaining information from them* about the emergency. Older proposals for emergency management have overlooked developing a step-wise and methodical consideration of the phases within an emergency situation. Keeping this modular design goal in mind, I present a solution to enable various administrative domains, varying *scales* of emergencies and *proactive* real-time information gathering to respond better to situations. Current implementations of emergency situations are not dynamic enough to take into account real-time information (which is essentially a resource, in the abstract sense) that might be made available to the emergency responders, by contacting users on the ground. By delineating a resource-scheduling strategy into different *schedulers* or *agents* I demonstrate the utility of evaluating various response mechanisms, which will contribute to my overall understanding of how well systems perform, in times of crisis.

5.1.1 Overview

My contribution is in the design, implementation and evaluation of computer-mediated agents by adding agent-driven communication services and agent-driven dynamic visualization interactions into my software. The central communications server, relational database services and machine algorithm components did not need to change to support my experiments. I further added a layer to the system, comprised of cell phone calls in the area, which enables me to identify users co-located with the emergency. The overall system is shown in Fig. 5-1.

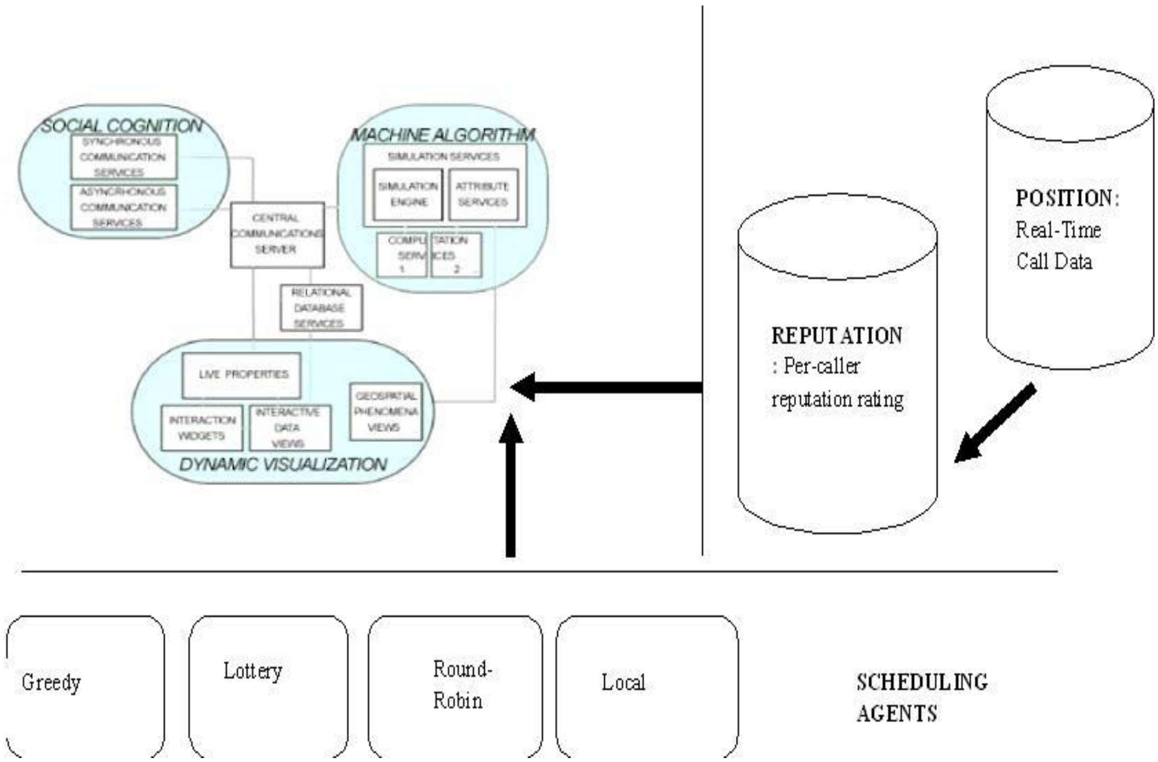
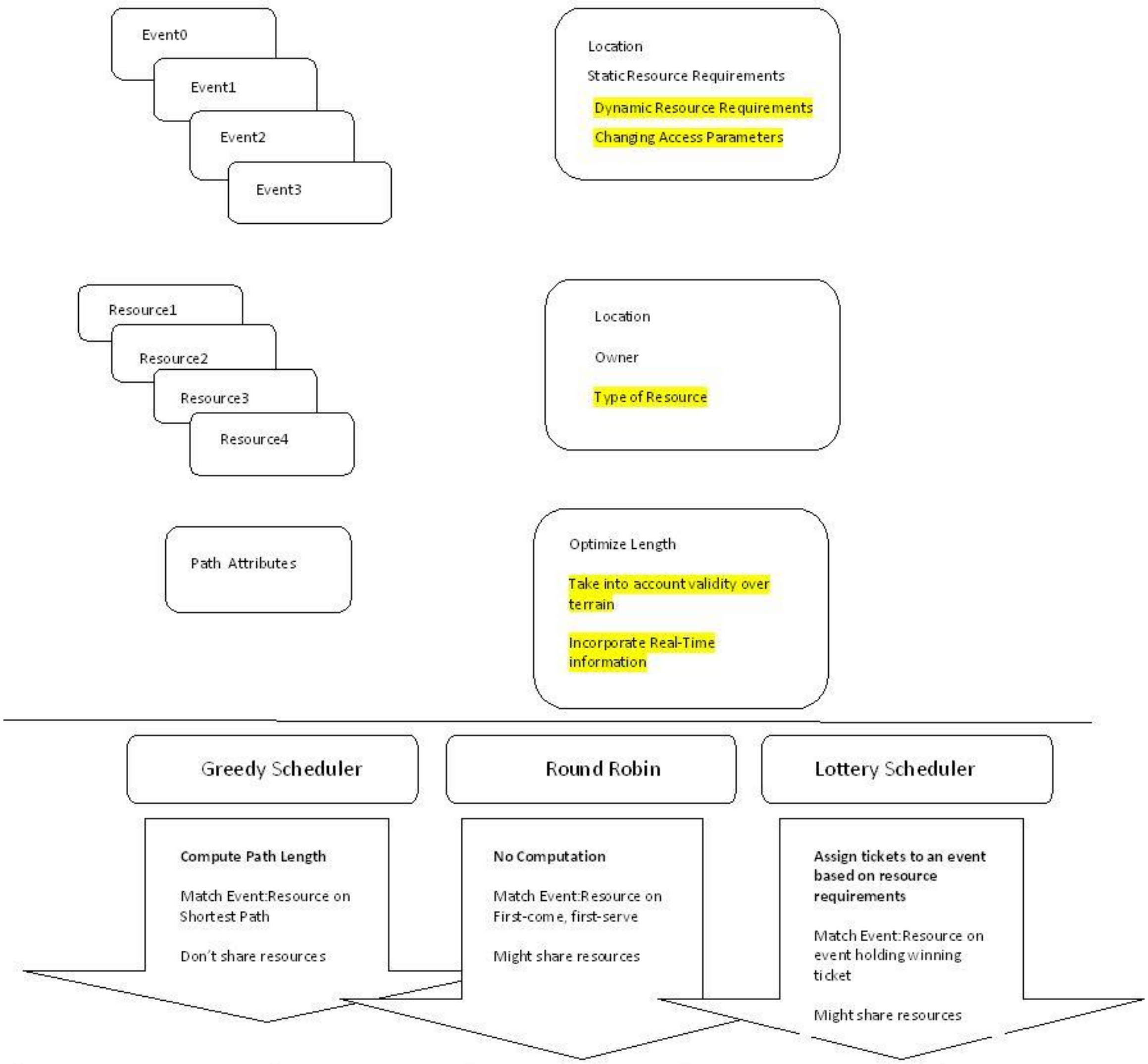


FIGURE 5-1. **System Overview.** My system includes data-sources that pertain to live calls, reputation metrics assigned to individual callers and the scheduling methodologies of three schedulers, Greedy, Round-Robin and Lottery.

5.1.2 System Architecture

Fig. 5-2 shows the overall system architecture of Reverse 111. The system has several modular entities that act in conjunction to simulate emergencies, resource-requirements of the emergency events, resource locations, path attributes considered while routing a resource to the requesting event, schedulers that are used to evaluate various metrics



Ease of RT Resource Requirement changes	Easy	Medium	Hard :
Ease of crossing administrative domains	Hard	Ineffective	Easy, issue tiered- tickets
Ease of Path-re-computation	Ineffective	Easy	Medium

FIGURE 5-2. **System Architecture – R-911.** My system includes data-sources that pertain to live calls, reputation metrics assigned to individual callers and the scheduling methodologies of three schedulers, Greedy, Round-Robin and Lottery.

including the ease of incorporating real-time information from users co-located with those emergencies, the ease of transcending administrative domains when responding to an emergency and the ease of path re-computation given that in a crisis situation the initial assumptions about the environment are not valid if the situation changes (for example, in the case of an earthquake, the roads might not be functional along the line of the fault and may vary in time as follow-on quakes happen or other constraints present themselves). I discuss the main entities of my system below with working examples:

EVENTS These refer to emergencies that are occurring in an area, that are represented in my system by means of icons appearing at certain latitude, longitude combinations over a geographic area. The icons chosen for these represent the resource requirements for these events. There are four different types of resources an event could need, and this is indicated by different colors. Figure 5-3 shows the iconic depiction of an event in my system.

FIGURE 5-3. **Event Depiction.** I show the iconic depiction of an event in this figure wherein each color stands for a different *type* of resource that an emergency event might require. Examples of these include fire-trucks, medical ambulances, police-vehicles or military support. In Reverse 111, an event does not have fixed resource requirements.

The attributes associated with each event are outlined below:

Event Location refers to the actual physical location of the event;

Event Static Resource Requirements describes the static resource requirements that are assigned upon the creation of every event;

Event Dynamic Resource Requirements describe the changing nature of an event with the progress of time. While I borrow from the RimSim Response depiction, I extend an event's resource requirements to be more *dynamic*. In the original RimSim Response, an event is static in its resource requirements and this is not a realistic depiction of any emergency, where the needs are constantly changing over time. Additionally, the objective of Reverse 111 is not training (as was the case with RimSim Response), rather it is to make a case for obtaining better information from citizens or cell-phone users that are co-located with an emergency. Therefore, my system needs to model an emergency event more realistically in order to support end-to-end evaluation;

Event Changing Access Parameters describes how the areas surrounding an event might change over time, given that my initial assumptions about an emergency may not be valid as time goes on.

RESOURCES These refer to the resources that are applied in order to ameliorate the emergencies. Examples include various state services and every resource has a number of attributes detailed below:

Resource Location refers to the geographical location of a resource. For example, a fire-truck might be at a fire-station when the emergency event occurs or might be returning to its home-base and therefore at a different location at the absolute time-stamp when its required.

Resource State indicates whether the resource is stationary or is in motion towards the event of interest.

Resource Owner refers to which administrative domain owns the resource. These could include state bodies such as the fire-department, police, military and can also be extended to include *commercial* emergency responders, which motivates my lottery-scheduler to evaluate a heterogeneity in such resource ownership.

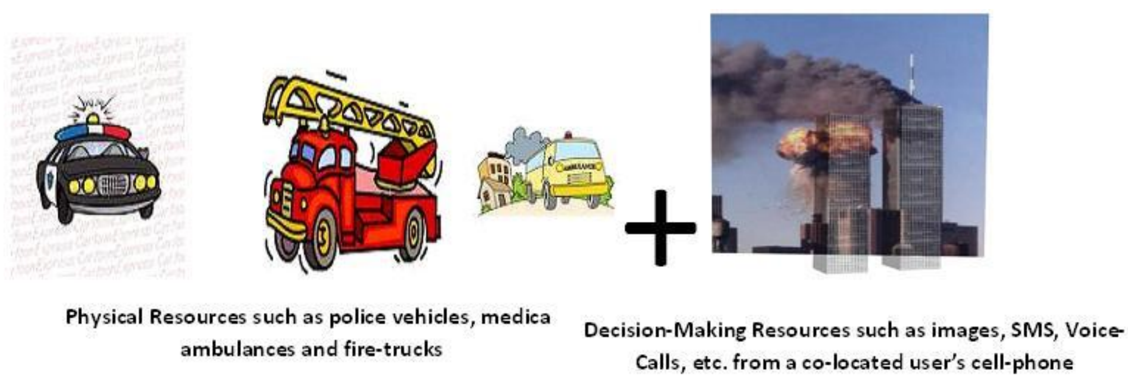


FIGURE 5-4. **Resource Depiction.** Physical resources are static and decision-making resources are more dynamic in my system.

Resource Type refers to whether the resource is a static resource (such as the ones described before) or a more dynamic resource, such as a user providing images of the façade of the building that has fallen down. The latter is more information than an actual physical resource and can be used for decision making. This is separate from the physical resources but an integral component of Reverse 111. Fig. 5-4 shows the types of resource used in Reverse 111.

PATH ATTRIBUTES These refer to the set of allowable paths over the map, while trying to get the physical resources to the emergency event.

Path Length My system assigns path-length depending on the distance of the resource from the emergency event. In the case where my system evaluates various scheduling policies, these path lengths may be pre-computed, based on requirements.

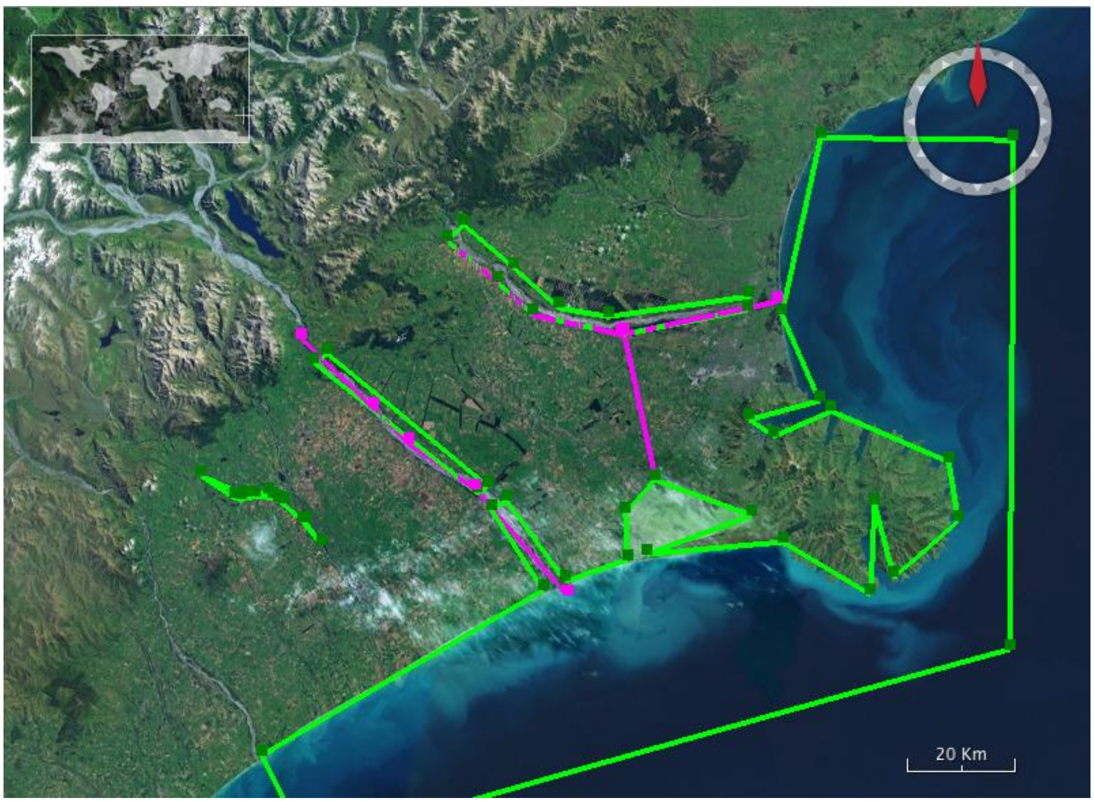


FIGURE 5-5.**Valid Regions** The valid paths that a resource can traverse to get to an event are based on demarcated areas, such that the resource is not trying to cross the sea, while trying to reach an event on the other side of the administrative or geographic divide.

Path Validity a valid assumption during the occurrence of a crisis is that all roads are not equal and all paths may not be open for the resource to take, in order to get to the event. For example, a bridge might have collapsed or an area might be too flodded to navigate. In these cases, the path validity is computed as a function of (valid_flag, time_slot) wherein the valid_flag indicates whether or not a path is valid and the time_slot indicates how *long* the flag contains robust information.

Path re-construction based on live information This is based on re-constructing the path (including length and validity) based on live information provided by co-located callers. Fig. 5-5 shows the valid regions being marked in my system such that the ocean is marked invalid on a resource-path towards the event.

5.2 Resource Schedulers

In order to simulate which emergencies are dealt with first, within a response scenario, I a number of agent services to the dynamic visualization component of RimSim Response. The three different schedulers I designed include the ***greedy*** scheduler, the ***lottery*** scheduler and the ***round-robin*** scheduler. The performance of these schedulers are compared and contrasted in the coming sections, for reference. I also add another service called the ***oracle player*** that controls the progress of the system, to run an entire scenario of multiple incidents occurring and being responded to, and run these schedulers in turn, in order to understand the differences in the optimality of various approaches to solving the same problem. I propose augmenting the schedulers with live call data, obtained from understanding a callers *location* and *reputation* and informing the schedulers decision dynamically with the incoming information.

Three parameters differentiate agents from one another including their policy on sharing resources with neighbouring jurisdictions, their strategy for determining which active unresolved incident should be attended to next, and their strategy for determining which available resource to assign the targeted incident. Since administrative boundaries exist in real-life, the performance of these agents mirrors how a real first-responder might think. In the case of a ***greedy*** scheduler, the attitude being simulated is one of the mayor of Christchurch always solving his or her local incidents first with their resources and never sharing anything. In the case of a ***round robin*** scheduler, it mirrors solving incidents one after another, in their order of appearance, without considering the class, severity or scale of the emergency such that all incidents are given equal priority. In the case of a ***lottery*** scheduler, this could mirror a commercial emergency responder that has to await their turn at state-based resources (hence the casting of a lottery by the oracle, in my case representing the state) before arriving to solve the incident.

5.2.1 Greedy Scheduler

The greedy scheduler allocates resources to incidents based on the most greedy approach, that of pre-computing resource distance, storing the shortest path and allocating resources to events based on which resources are closest, even if the events are serviced out of order in terms of arrival time.

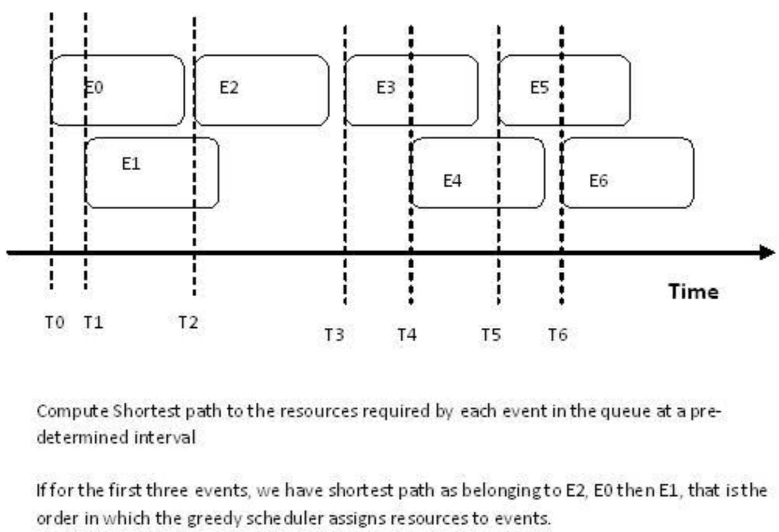


FIGURE 5-6.**Greedy Scheduler** This scheduler pre-computes the shortest path of resources to events and assigns resources based on the closes event.

Figure 5-6 shows a greedy scheduler wherein the system sees the arrival of emergency events E0, E1, E2,E3,E4,E5 and E6 at time T0, T1, T2, T3, T4, T5 and T6. The scheduler then pre-computes the shortest path of the resources required by each event, the pre-computation happening at a set time-interval, such that in case event E2’s resources are the closest to it, followed by E0’s resources and then E1’s resources, that is the order in which resources are dispatched to events. This scheduler is also administratively greedy in that it does not share its resources across administrative domains. In case an event's resource requirements change dynamically, it is easy to incorporate this change within the greedy scheduler as there is a pre-computation of shortest path of resources to events and this is dynamic in nature. In case the administrative domains need to be transcended (assume a resource is right at the border of two counties wherein the event the resource is closest to is not in the greedy scheduler's administrative domain, rather the one next-door) this scheduler will not perform well as it doesn't share resources. In case the path needs to be re-computed, this scheduler is again ineffective as the pre-computation of shortest path will undergo several extra steps to calculate the new shortest path and cannot use cached values.

5.2.2 Lottery Scheduler

In case an emergency event has to be handled by one or more commercial service providers, there might be contention for the cirtical resources (such as fire-trucks etc.) that are not essentially in the private service providers domain of resources. In this case, events could get *starved* of resources if they wait for the authorities to make up their minds as to who to dispatch next. To handle this scenario, I propose a lottery scheduler [waldspurger], which borrows ideas from the Mach Kernel, to implement *proportional fair-share scheduling* and adapt the same to Emergency Management.

Figure 5-7 shows a lottery scheduler wherein the system sees the arrival of emergency events E0, E1, E2,E3,E4,E5 and E6 at time T0, T1, T2, T3, T4, T5 and T6. This scheduler uses the concept of currencies to denominate tickets to each of the events. Each currency (resource-type) is *funded* by tickets at a certain exchange rate (i.e, a military-truck might be more expensive to dispatch than a fire-truck). Every event is assigned tickets from a pre-set pool of tickets depending on their resource requirements. A random-number generator picks the winning ticket and the list of events is traversed until the event holding the winning ticket-number is found. This method avoids starvation amongst the various events and also designates tickets based on resource requirements therefore giving every event a fair chance at obtaining the resources they need.

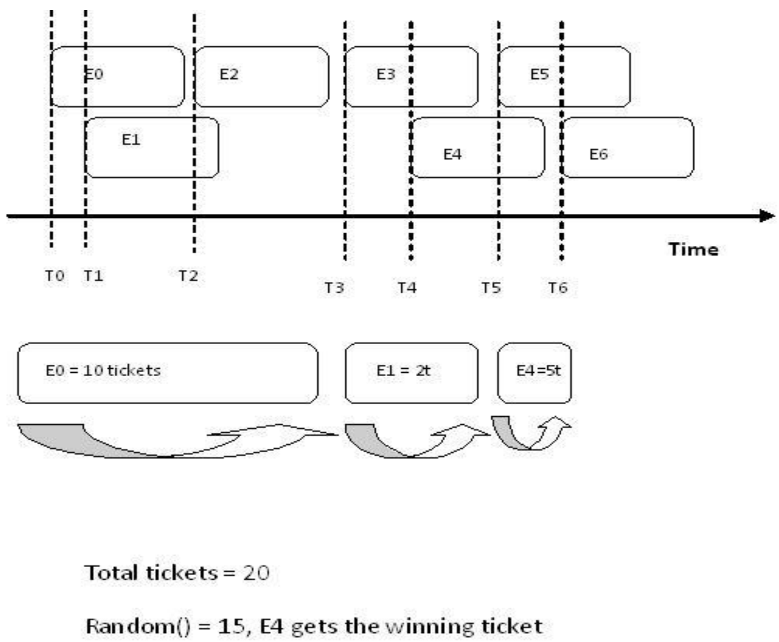


FIGURE 5-7.**Lottery Scheduler** This scheduler pre-computes the shortest path of resources to events and assigns resources based on the closes event.

Since the tickets are assigned before hand, in case an events resource requirements change in real time, this will affect the performance of the scheduler as it has to recomputed tickets on the fly, which makes the subsequent steps of casting the lottery through the random-number generator and then finding the winner stall, until the requirements are re-calculated. Crossing administrative domains with this scheduler is easy as tickets can be tiered or valued differently based on the notion of different *currencies* for different resources. In case the path needs to be re-computed, this scheduler's performance will not be affected that much as the only event affected will be the current event and the others can have their paths re-assigned in the background, while the current set of events are being handled.

5.2.3 Round Robin Scheduler

Figure 5-8 shows a round-robin scheduler wherein the system sees the arrival of emergency events E0, E1, E2,E3,E4,E5 and E6 at time T0, T1, T2, T3, T4, T5 and T6. This scheduler uses a first-come, first-served policy wherein the order of appearance of events determines which event is allocated resources.

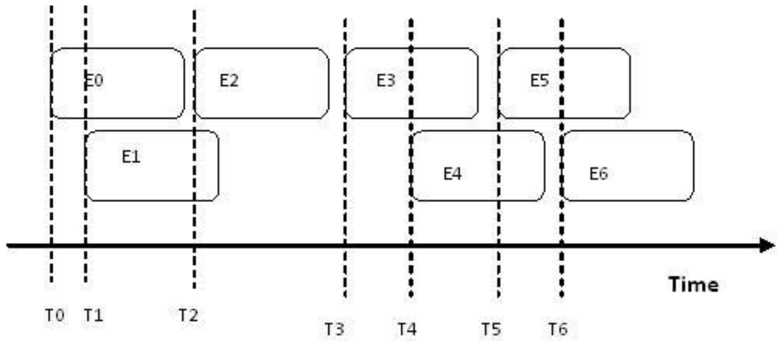


FIGURE 5-8 **Round Robin Scheduler** This scheduler is the most primitive of all three

5.3 Visualization Interface

In this section I briefly present the visualization interface, which was built to simulate emergencies in New Zealand. I focus my discussion to the South Island for purposes of example and this tool is built to incorporate really any region in the world. There are three steps in the visualization, *setup and initialization*, which allows the user to set the administrative boundaries and invalid paths on the map, *visualizing events* which occur on the map, depending on the type of the emergency, *visualizing resources* which are placed at certain locations and are of four primitive fixed types (red, blue, green, yellow) for purposes of this example and *visualizing co-located callers* that provide dynamic, real-time information about the emergency.

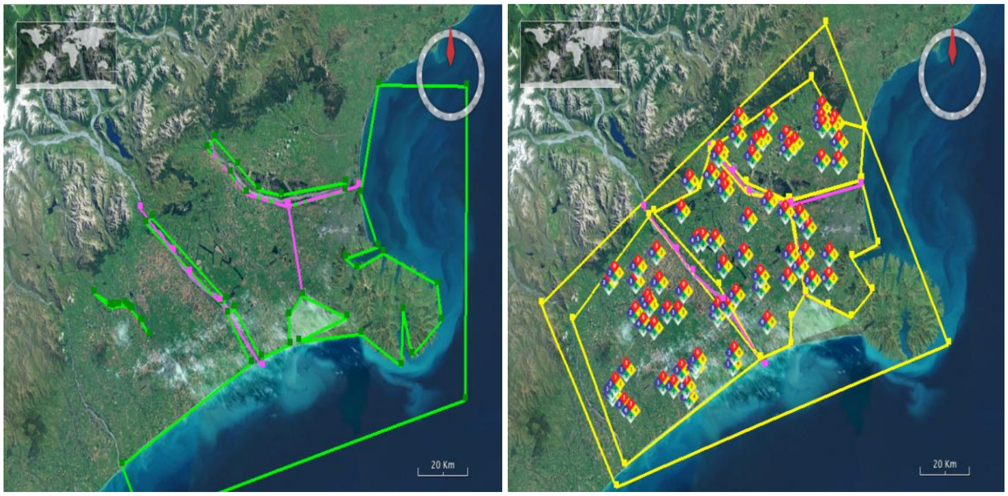


FIGURE 5-9.**System Setup** Several parameters are initialized in this step including setting the administrative domains (shown in pink), setting invalid paths (shown in green), and setting boundaries for each domain (shown in yellow).

Figure 5-9 shows the system setup phase of this visualization tool. Parameters that are initialized during setup include setting administrative domains, shown in pink on the panel on the left-hand side and setting invalid paths, shown in green such that the ocean, the gorges (Ashley, Rakaia, etc.) are not valid paths for resources to travel through while getting to the event. Further, the system sets the boundaries for each administrative domain to solve events in. In this example, the global domain is the outermost yellow-line enclosed area on the panel on the right hand side. Each of the smaller yellow-line enclosed regions correspond to the administrative regions of Christchurch, Ashburton, Oxford and Rangiora. Figure 5-10 shows the visualization of randomly placed events, with different resource requirements at time T1 and time T2 (such that T2>T1), within Christchurch. Figure 5-11 shows the visualization of events that are related to a bio-weapon attach, an earthquake and a tsunami in Seattle. The placement of events in Figure 5-11 is more methodical and relates to the actual spread of the event. These scenarios are described in detail below:

1. Bioweapon scenario: Incidents are clustered closely together, with a spatial distribution approximating a Gaussian distribution with a 1-mile half-width. Because the incidents are so closely set, two of the agents have no incidents in their regions at all. The incidents are timed to follow a slightly smoothed Poisson distribution with $X=1$ in units of 15 seconds, so most of the incidents activate immediately, the rate of incident activation dropping off quickly until the last incident appears 48 seconds from the start of the simulation.

2. Earthquake scenario: Incidents are distributed uniformly over the playing field - representing a shallow source of disturbance on unstable soil - so that each region has a number of incidents proportional to its geographic size. The temporal distribution is the same as in the bioweapon case.

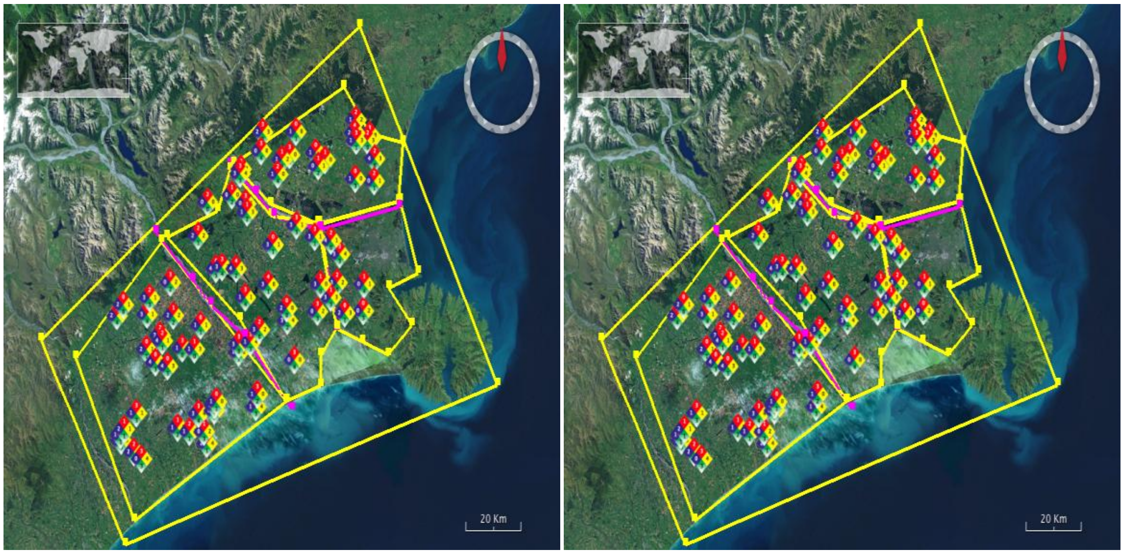


FIGURE 5-10 **Visualizing Events** Events can occur in random order and in this figure I show the appearance and spreading of events around the South Island of New Zealand as the first example.

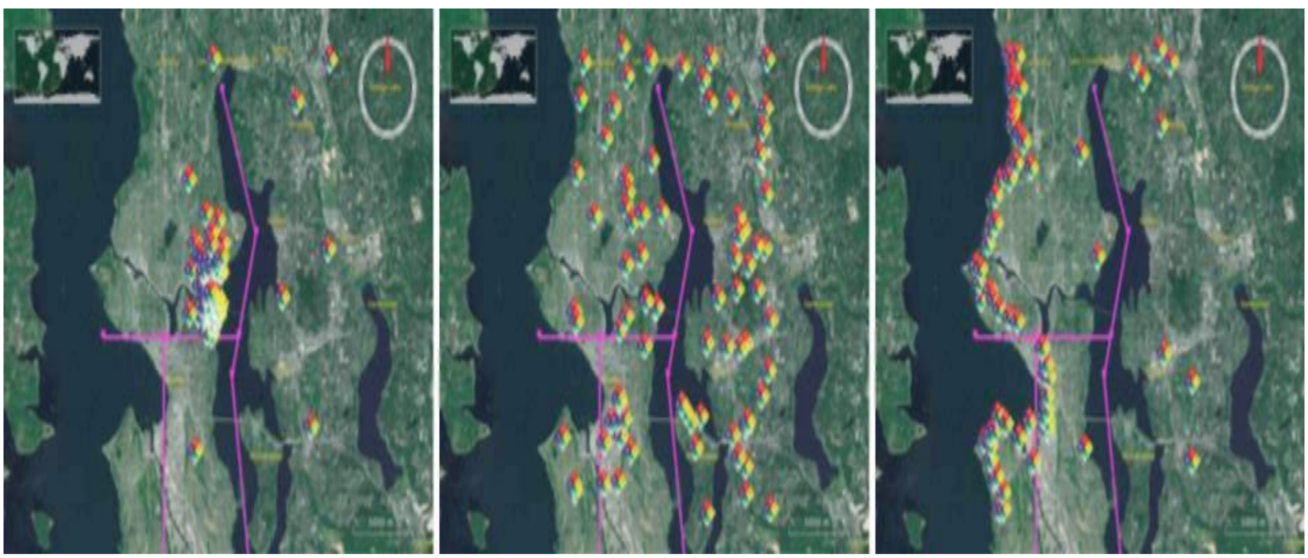


FIGURE 5-11 **Visualizing Events** Events for a bio-weapon attack, an earthquake and a tsunami in Seattle are shown in this figure.

3. Tsunami scenario: Incidents are placed by hand along the shoreline of Puget Sound but with the same number of incidents as the earthquake scenario. The temporal distribution of incidents is the same as in the bioweapon and earthquake cases.

   Figure 5-12 shows the visualization of static resources such that there are four types of resources placed at different locations within the administrative domain.

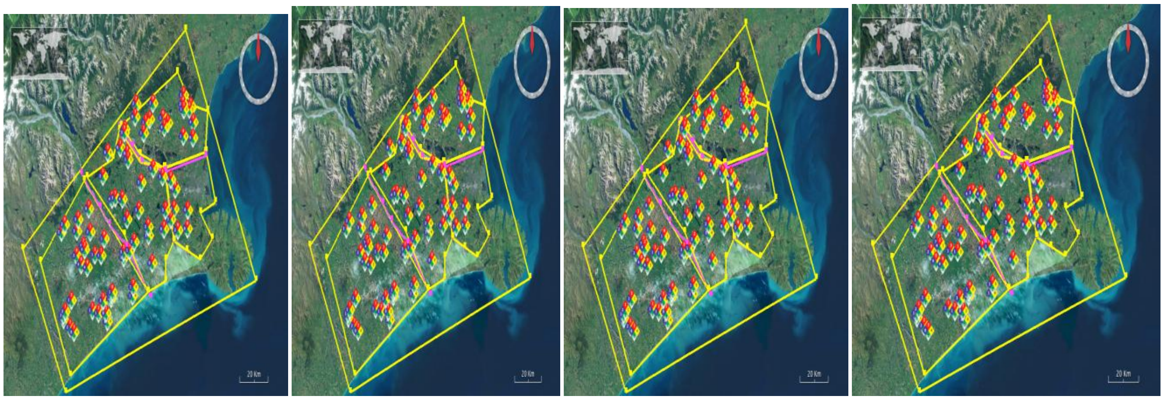


FIGURE 5-12 **Visualizing Resources** Static resources are placed at random locations within the city.

Visualizing co-located cell-phone users is shown in Figure 5-13 such that different number of users are present at different locations at different points in time. Figure 5-14 shows the movement of resources towards incidents.

5.4 Performance Evaluation

In this section I present my preliminary evaluation results for throughput and latency of the three different schedulers. In my set-up an incident occurs every 4 seconds and the greedy and lottery schedulers perform their pre-computation simultaneously with incident arrival, thus there is no lead time. I present my results over 600 incidents that are occurring in the South Island of New Zealand, at random locations.

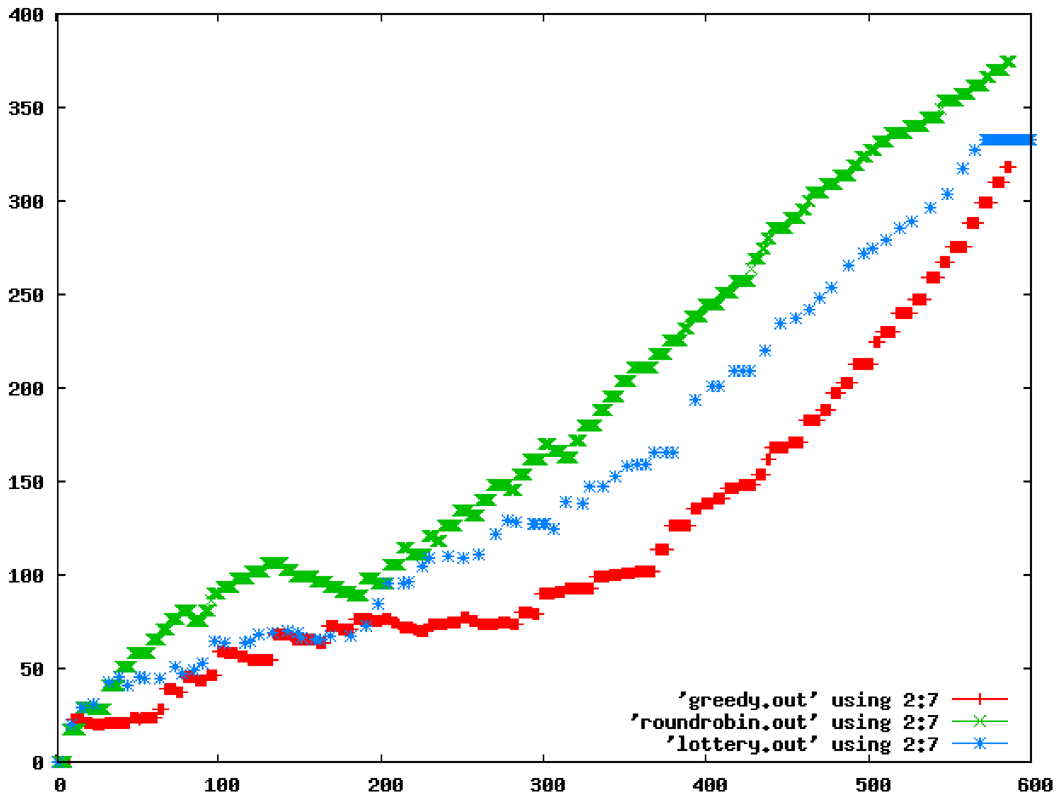


FIGURE 5-15 **Scheduler Latency** The x-axis shows the number of incidents that are occurring and the y-axis shows the average latency in the scheduler assigning resources to events.

Figure 5-15 plots the average latency between the event occurring and having its resource-requirements satisfied. I note that the greedy approach out-performs the round robin and lottery schedulers by quite a bit upto 200 incidents, at this point, the three approaches seem to perform equally. This is followed by the greedy scheduler out-performing the round-robin and the lottery schedulers at a steady rate. At the heaviest event load, the greedy scheduler is closer in performance to the lottery but both out perform the round robin approach by quite a bit. The round robin scheduler is the most primitive approach to resource scheduling. This does not take into account the proximity of the resources to the events or the resource requirements of individual events. The lottery scheduler is proportional fair-share but involves allocating and de-allocating resources to events as the winning ticket changes. For a small number of incidents, the lottery scheduler performs comparably to the greedy approach, at one point even doing better (180 incidents), however the greedy approach scales better.

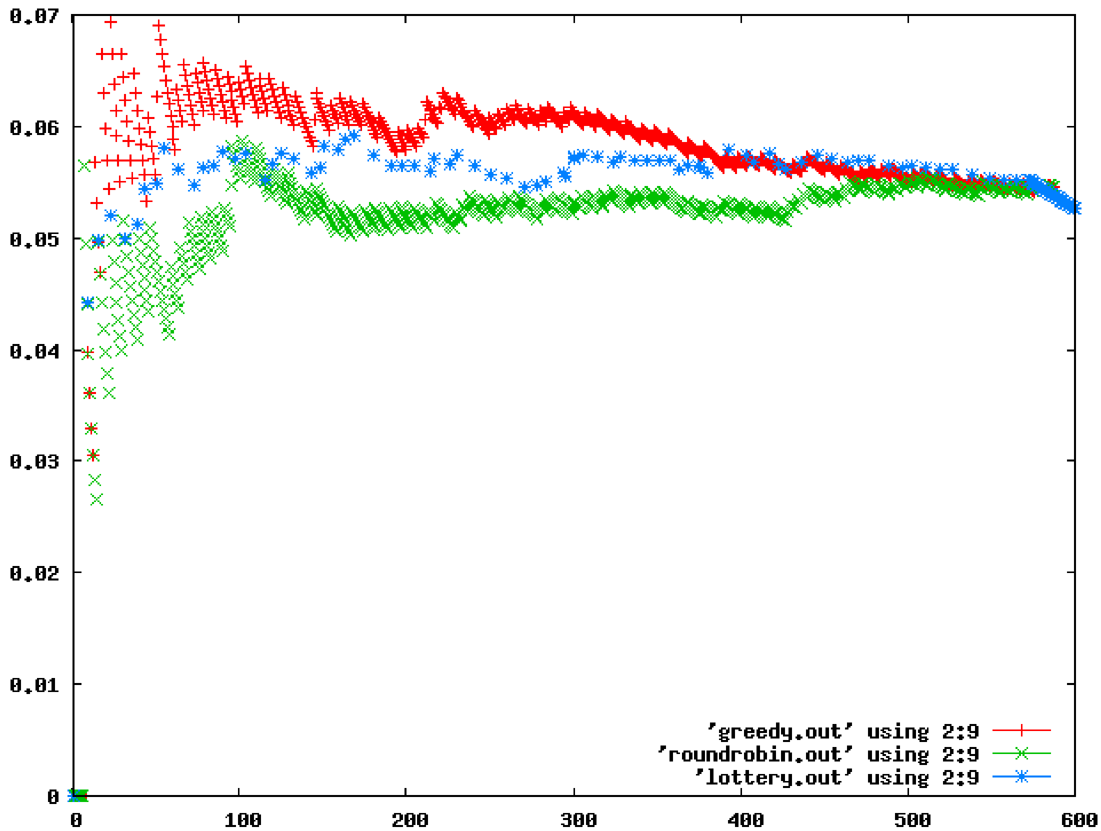


FIGURE 5-16 **Scheduler Throughput** The x-axis shows the number of incidents that are occurring and the y-axis shows the average throughput in the scheduler assigning resources to events.

Figure 5-15 plots the average throughput of each of the three schedulers. I note that the greedy approach out-performs the round robin and lottery schedulers by quite a bit upto 100 incidents. Beyond this, the throughput (number of events handled by each scheduler within a certain time interval) deteriorates as the number of events increase with the greedy scheduler outperforming the lottery scheduler and the round-robin scheduler, in that order. This result is best explained as the round robin scheduler has to wait till the event the resources are currently serving, free up, before they are re-assigned to other events. In the case of the greedy approach, pre-computing the shortest path also takes into account the final position of the resource (i.e, if the resource moves to handle and event, that position is its final position) and therefore when the shortest path is re-computed once that resource is free, it is still the shortest path to the next event it will service. The throughput of the three approaches gradually converge as the number of incidents increase demonstrating the need to evaluate these proposals at different granularities of event occurrence.

5.5 Related Work

The Pacific Rim Visualization and Analytics Center coordinated at the University of Washington, known as the PARVAC to other regional centers, has been working on developing improved shared artifacts for communicating during community-wide first response activities. Acknowledging the magnitude of the cognitive load involved in maintaining situation awareness for effective first responder behavior they have studied successful distributed cognition methods for embedding mission-critical knowledge in shared visualization tools. Their primary focus has been on the embodied mind model that suggests humans need a wide arrange of external thinking tools in order to embody an internal model for effective action. In order to share situation awareness, they have explored the emerging model for augmented cognition – a model that looks to improve social cognition between humans given solid shared artifacts and associated computation support.

5.6 Conclusions

In this chapter, I introduce different types of schedulers to assign resources to emergencies, each scheduler differing in the policy it adapts to resolve contention or pick emergencies from a list of emergencies. One of the main reasons to understand how various scheduling policies perform is that several actors need to co-operate in order to mitigate emergencies. These actors include, but are not limited to, various first-responder units, citizens groups, emergency care providers etc. When there is a multitude of administrative domains to be traversed both in policy and with respect to resources, understanding the performance of schedulers gives us some insight into how these various actors might end up co-operating and performing in real situations. For example, the lottery scheduler could well be applied to the case where several first-responder units have resources to offer and differing numbers of resources. In order to determine which responder is allowed to assign their resources to an emergency, some authoritative source (in our case, the oracle player) might have to cast a lottery ticket to the responder. Thus, our schedulers try to mirror real-time discrepancies in the number of actors and resources they offer to the emergency situation and try to evaluate the performance of the same.

6 Manikarnika

This section proposes a framework for emergency management, wherein the entire continuum of an emergency situation is assessed and taken into consideration. An emergency situation can be thought to have multiple distinct phases from occurrence, collection of information or data (either before or after occurrence), prevention and response. Depending on the phase that a common citizen or a first responder finds themselves in, the metrics that determine performance and success of solutions to address them, will vary. Emergencies vary in scale and severity and depending on the actual location of the emergency, both the impact of the emergency and the response that is enabled, affect the quality of managing the situation. In this study, I propose a framework for emergency management that discusses various contributing factors including metrics, data, tools and technologies that need analysis in order to determine the efficacy of the solution and propose a proactive, rather than reactive solution to the problem. Section 6.1 presents the framework, Section 6.2 discusses the features of cell-phones that render them useful, Section 6.3 shows the application of AR to provide visual cues to the civilan that is assisting an Emergency Centre with information, Section 6.4 discusses related work and Section 6.5 presents my conclusions

6.1 A Framework for Emergency Management

The management of emergencies has necessitated a wide field of research into warning systems, first response mechanisms and effective mitigation of the disasters. The taxonomy of emergencies can vary depending on the severity, scale and nature of the emergency. Examples include civil emergencies, medical emergencies, natural hazards, epidemics and large-scale terrorist attacks. Depending on the nature of the emergency, the tools that are used to manage the emergency vary. I argue that the most ubiquitous sensor that can be used across the continuum of emergency detection, warning, response and mitigation is the common cell phone, simply because of its proliferation and availability. In this study, I present an array of sensors that have been proposed in literature for use across a variety of emergency classes and present my case for using the cell phone in most cases as a good sensor for emergency situations. The features of the cell phone render that it a viable sensor include providing location information, being a highly prolific device and being a two-way communication channel where both the user can call in for help and emergency managers can contact civilians for information. The telephone network of a leading services provider in New Zealand is profiled for calls made on mobile phones in order to establish a set of macro and micro level information about various entities including users (the population carrying the phone), fixed resources (such as cell phone towers, which are the backbone of communication) and traffic density across the geographical expanse of a city on normal days and the days of special events. My study also motivates the need for varying levels of detail in the information established in order to enable fast and efficient emergency management.

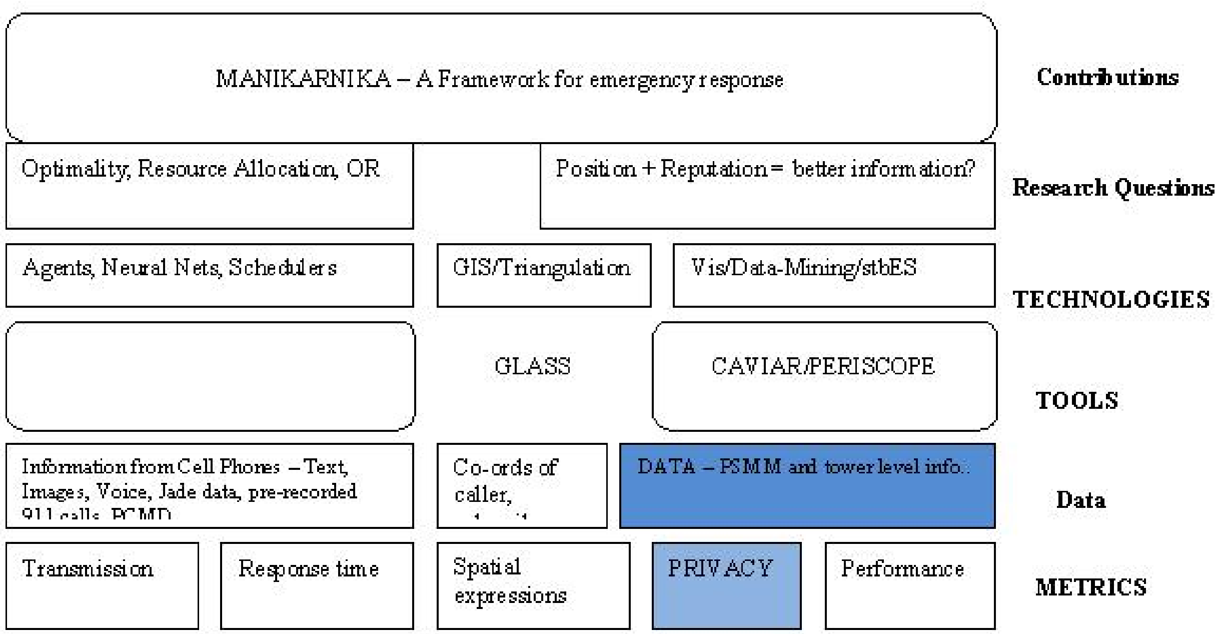


FIGURE 6-1. **Manikarnika Framework.** I present a tiered approach to understanding the elements of the framework including metrics, data-sets, tools, technologies and research questions, which contribute in tandem.

I propose a framework for emergency response that takes a tiered approach to problem solving and exposes the modules involved in a step-wise fashion, shown in Fig. 6-1. At the first level, I have the metrics that are used to determine the effectiveness of any proposal that might want to mitigate emergencies. For example, an emergency in India is very different from an emergency in Seattle to an emergency in New Zealand. In the case of India, the effectiveness of the solution might depend on the number of lives the solution or the response actions are able to save (as this is a very densely populated region of the world) versus in Seattle, the efficiency of a solution might involve ameliorating the overheads occurred by cross-county and administrative co-ordination (as legislative procedures in the United States are very stringent) versus New Zealand, where the metric to be considered might be the sheer distance traveled in order to get the resource to the right place at the right time to mitigate a disaster or respond to an emergency. In the second tier, I have the data that might be used to collect information or traces at the various levels. In my work, I am heavily reliant on per call measurement data, which is essentially information from cellular phones including not just the co-ordinates of the phone as recorded by the network but also the information the crowd is able to send back to the emergency responder. Examples of this might include citizens sending text messages and images, voice calls received pre-recorded 911 calls, etc. Location of the users is recorded by the network and provides the co-ordinates of the caller. Other data such as the location of the tower itself might be relevant in the most pathological case where a cell phone tower falls down and its neighbors need to be activated in order to handle the load and migrate callers to functioning towers. In the next level, I discuss the variety of tools one might use to enable these solutions including GLASS, which is a geo-location tool used to find out the location of the caller, CAVIAR/PERISCOPE which are visualization tools developed to understand where the towers are located, how users are moving between these towers, how far they move during the call, etc. I also have REVERSE-111, which is essentially a proposal for proactive crowd sourcing which alludes to calling a user on the ground, that is co-located with an emergency or a disaster and obtaining information from them on the severity, nature and scale of the emergency. The scientific technologies that contribute to the larger understanding of the underlying mechanics of the problem at hand include the use of agents, neural nets and schedulers that have found more conventional uses in the research areas of reputation management, AI and operating systems design. Further, there is the field of location algorithms that use a combination of GPS-enabled methods and triangulation to obtain the exact location of a user, at various resolutions. Tools used to visualize the emergencies are also paramount to giving the responders meaningful interfaces to react properly to such situations. Technologies such as stbES have changed the way in which AR can be enabled on hand-helds and these are relevant as demonstrated in a later section, with examples. The bigger research questions where I draw a lot of the core work from comes from the areas of Optimization, Resource Allocation and core Operations Research and added to this I am trying to answer the question whether a combination of position and reputation leads to better information about situations, which could extend to both the commercial and non-commercial sphere.

6.1.1 End-to-End Design

The continuum of emergencies starts with the occurrence of the emergency or early warning systems (depending on predictors being set up), the assimilation of information and data pertaining to various = aspects of the emergency, the prevention of the emergency if possible and responding to the emergency. Depending on the nature of the emergency, a variety of tools and techniques might render themselves relevant. In this study, I present a comprehensive analysis of sensors that are used in emergency situations wherein the scale, severity and nature of the emergencies might vary greatly. The emergencies themselves could belong to a wide category including natural hazards, civil emergencies, medical emergencies, epidemics and large-scale terrorist attacks. The parts of the continuum in which the sensors are used are also varied. For example, certain sensors are used to detect the occurrence of an emergency. Other sensors might be used to assimilate data and information about the emergency, requiring them to have a higher hardware specification, when compared to simple sensors. Sensors can also be used to simulate emergencies, issue early warnings to people affected by the emergency and mitigate emergencies. Fig.6-2shows the continuum of an emergency situation.

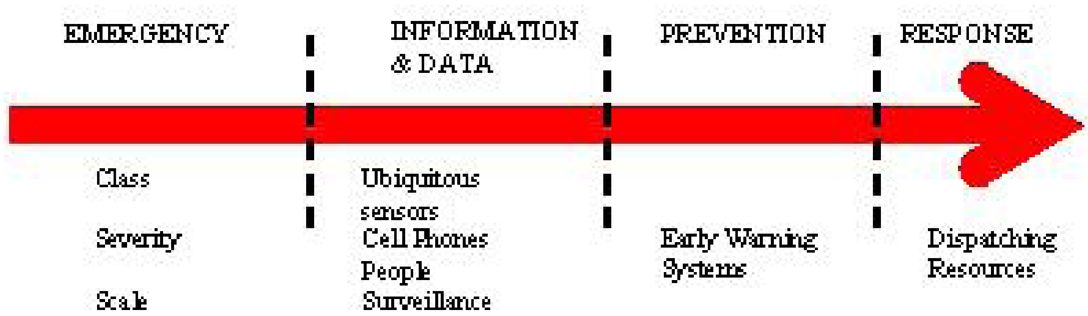


FIGURE 6-2. **Emergency Continuum.** Different phases exist within emergency management including occurrence, surveillance, prevention and response.

6.2 Cell Phones as ubiquitous sensors

Three main features act together to present the cell phone as a viable alternative for emergencies:

6.2.1 Location Information

Cell phones can be used to pin-point the exact location of users across a region [16] and in the case where this mode of communication is still viable, this is a very important piece of information to mitigate the emergency. In the case of large-scale natural disasters, I might encounter the case where the cell-phone towers themselves are no longer available, rendering a sizeable coverage hole in the network. In this case, the most proximate communication region is important to establish to both gather information from users that are able to communicate back with the emergency management team about what they see in terms of the disaster and to possibly evacuate other population groups in the vicinity. In cases such as the co-ordinated terrorist attacks in Mumbai and New York, the cell phones become very relevant as reliable sensors. In both cases, a series of attacks were co-ordinated and carried out in tandem. The scale in these two situations is different as Mumbai is far more populated than New York. However, the basic utility of the cell phone remains indisputable. To be more specific, in the case of the Mumbai attacks, the chronology of events that unfolded makes a very good case for my work, as shown in Fig. 6-3. At 8pm on 26th November, 10 Urdu speaking men in inflatable boats came ashore in Mumbai. At 9:20 pm, two attackers opened fire at the Chattrapati Shivaji Terminus (I). At the Leopold cafe in Colaba, 10 more people were killed (E). At 10:40 pm, bombs went off in taxis in the Vile Parle regions (G). Two hotels, the Taj and the Oberoi (B) were attacked on the same day as the gunmen took people hostage. Fig. 6-3. Points of interest in the Mumbai Terrorist attack.

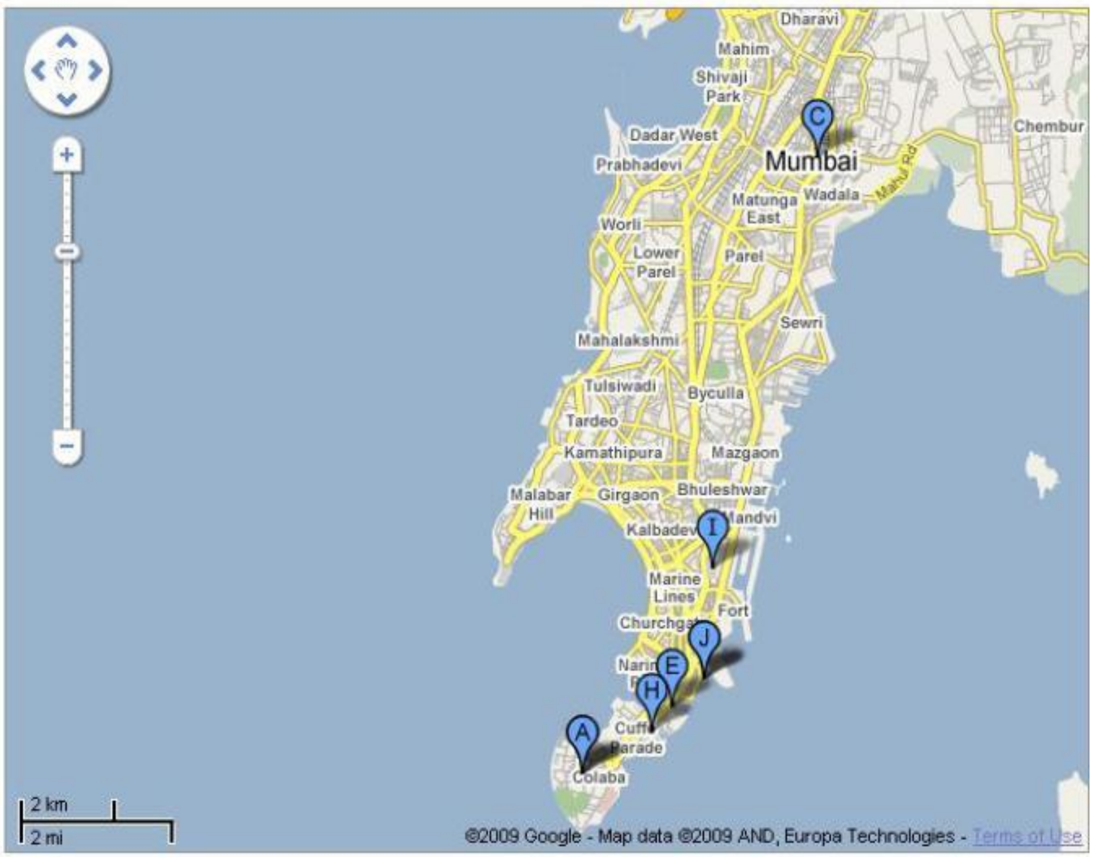


FIGURE 6-3. **Mumbai 26/11.** Points of attacks on November 26th, 2008 in the Mumbai blasts

November 26th, 2008 was a weekday (Wednesday) and the typical rush hours between 5-8pm, however, given the population density in India, any time during the day, the number of people using trains here is quite significant. The time at which the attacks were carried out are also important as with cell phone data, one can predict the times at which certain sections of the city are going to be most crowded and create contingency plans around that information.

6.2.2 Proliferation

Cell phones are in use all around the world, from the poorest to the richest countries. According to the world facts book, 60 percent of the human population carries or uses cell phones. The use of this technology is prolific even though the actual devices may be basic. The percentage of mobile subscribers is growing the fastest in Asian countries and increasingly, households prefer to have only cell phones, rather than having both a landline and a cell phone. The average price of a handset is upwards of 30 US Dollars or 1200 Indian Rupees, making it accessible to a vast majority of the population. Fig. 6-4 shows the number of cell phones sold in comparison to more sophisticated computing substrates such as smart phones and laptops. Even if I assume that most cell phones are basic models, the average number of transistors in the cards used in these devices is substantial. Extensibility of the cell phone as a device that can be suited for emergency management would not be too much of a stretch, if I identify the requisite characteristics before hand. I propose some ideas to use the cell phone effectively in the following sections, including expanding on the Reverse 911 idea in order to contact civilians on the ground, not just for early warning but also for information and data gathering.

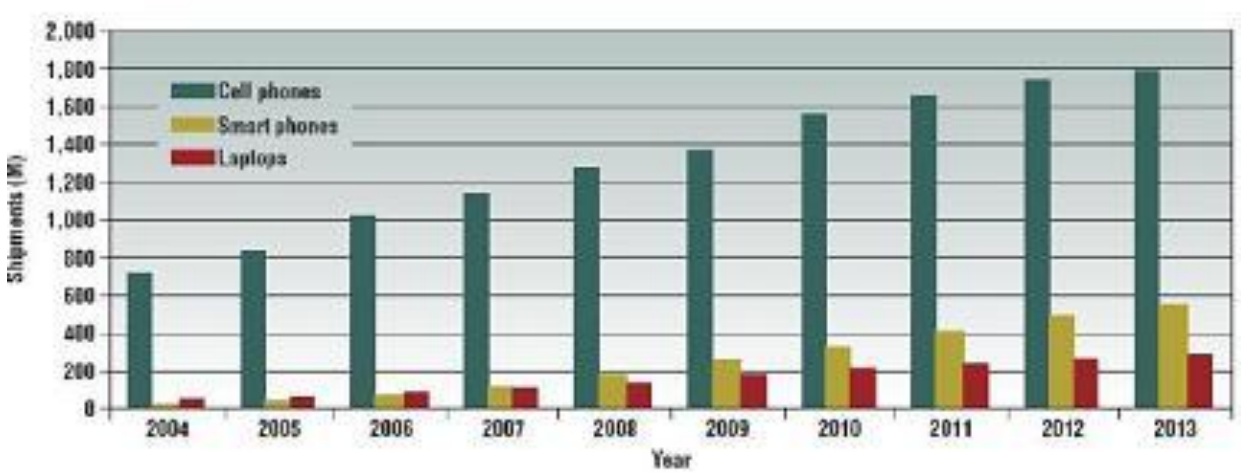


FIGURE 6-4. **Smart Phones**. Proliferation of SmartPhones, the most advanced and capable handhelds

6.2.3 Two-Way communication channel

The most important feature of a cell phone that renders it as a useful sensor is the fact that it is a two-way communication channel. While this statement may sound obvious, my applications are proposing the use of civilians on the ground as reliable sources of information and data pertaining to various emergencies. In older proposals, citizens have been contacted via cell-broadcasting or telephone messages from Emergency Management centers in order to get them out of a potential problem area. While the nature of the emergencies addressed by these solutions remains the same as ours, in that they cannot be predicted beforehand, my solution proposes that the emergency management center contacts civilians for a different purpose than simply early warning or evacuation. For example, if I knew the location of a user at the time of a potential terrorist attack, I might consider contacting them, in order to assimilate more information regarding the situation. Especially in the case of co-ordinated attacks, this proposal gains more reputation as the location of victims is crucial and communicating with them is just as crucial. Fig 6-5 shows the various metrics that can be obtained by means of caller records that use cell phones and how they relate to the emergency continuum. The information is divided into macro and micro levels and traffic characteristics. The parameters that are of interest in emergency management could be any of cell tower location, user location, cell tower aggregate statistics, and user density, varying levels of detail and a comparison or signature of a regular day’s traffic, when compared to a busy day. The utility in the emergency continuum presented in Fig.6-2 might vary, depending on the parameter that I can obtain from the various information levels.

Information	Parameter	Utility in the Emergency Continuum
Macro Level	Cell Tower Location	Information/Data gathering, Prevention Systems
Micro Level	User Location	Warning; Information/Data gathering; Prevention; Response
	User Movement	Prevention; Response
	Cell tower statistics	Information/Data gathering; Prevention
	User Density	Information/Data gathering; Prevention; Response
	Varying levels of detail	Warning; Information/Data gathering; Prevention; Response
Traffic Characteristics	Regular days traffic vs. Traffic on days with special events	Information/Data Gathering; Prevention; Response

FIGURE 6-5. **Levels of Information**. Macro and Micro statistics pertaining to the network inform different parts of the Emergency Continuum

6.3 Overcoming Communication Barriers

In this section, I expand of the utility of Augmented Reality in Emergency Response situations, in order to provide efficient and expeditious information, while addressing language barriers that can be used to manage the emergency. This particular path is commercially viable today with the emergence on AR-based advertisements, etc. and I present a brief prototype that explains the relevance of this technology to Emergency Management. Linguistic and other communication barriers present themselves on a magnified scale in times of an emergency. The simplest aids to civilians assisting emergency managers to relay information from the ground would be visual. I explore this alternative by means of proposing an extension of Reverse 111 that utilizes AR to provide visual cues to civilians and establish a two-way communication between the emergency response center and civilians co-located with the emergency, thereby presenting a case for AR in Emergency Response.

Fig. 6-6 shows the actors in the proposed system including the emergency response center, civilians co-located with the emergency and the data and information that flow between the various entities.

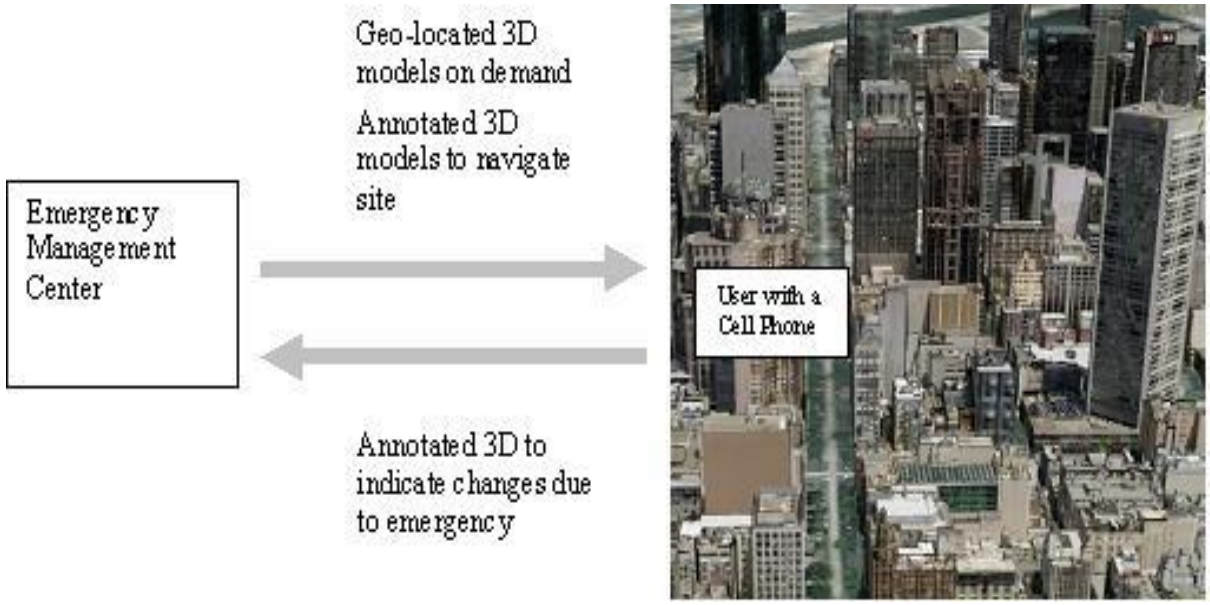


FIGURE 6-6. **Communication Barriers** . Visual cues enabled by such technologies as Augmented Reality on Phones can allow users that are assisting at the time of an emergency to receive visual cues and annotate the same to disseminate relevant information.

Fig. 6-7a-c propose the use of semi-transparent models that can be used in order to super-impose images or other content received from civilians that are co-located with the emergency with the old model of the city so that he emergency response center can best determine what resources to dispatch in order to effectively mitigate the emergency. In step 1, the Emergency Response center prunes the entire 3D model of the city (in my case NYC) to identify the section of the city relevant to the caller. 3D models of cities are obtainable from free websites such as Google Earth (where this one was obtained from) or can be created at varying resolutions and levels of detail as the city centers see fit.

FIGURE 6-7a. **3D Model of a city** .[Image Courtesy: Google]. The Emergency Response centre maintains 3D models of the entire city.

In the second step, the emergency response center sends the relevant parts of the model to the mobile phone user, co-located with the disaster.

FIGURE 6-7b. **Relevant images sent to user** In this step the Emergency Centre sends the relevant portions of the model to the user, in order for the user to annotate these models to provide information of what is happening on the ground.

In the third step, the user is able take an image of the *change* in the scene from their perspective as they have the 3D model to compare against. This image is sent back to the emergency response center for them to super-impose and update their 3D model with, post-emergency and dispatch resources suitably.

FIGURE 6-7c. **User-Annotation** In this step the user annotates the image to describe what has happened to the building in this case, for example annotating the face of the building that is collapsing, to send back to the Emergency Centre.

6.3.1 Extending Reverse 111 – Civilian Response

In my proposal, once an emergency occurs, an emergency management center typically contacts civilians that are co-located with the emergency, in order to obtain more information about the situation. In this proposal, I present the utility of 3D models of the city that can be sent to the civilians in order to:

(a) Have the civilian annotate where the disaster has occurred including such information as which face of the building has fallen down or which street has been blocked off by debris; and

(b) Have the emergency management center annotate the model they send the civilian with visual cues of where they would like the civilian to go in order to send them reports that are of more value (such as walk to the left of the building and send me a picture of the side of the building that faces King Street).

6.4 Related Work

The use of sensors for Emergency Medical care is probably the most wide-spread. Bluetooth compatible devices have been studied in great detail to assimilate data and information and interface to a central server by means of a cellular phone [120,121,122] . Sensors are also used in the case of emergency evacuations [123] from within buildings. Emergency response simulation using sensors have been performed [124] in order to co-ordinate and control the dispatch and use of resources required. Specialized computation of emergency exit routes in case of toxic gas attack [125] have been discussed where models are used to compute the dispersion of the gases. Here the sensors are those used by the live weather web service which is able to predict and supply weather patterns for any zip code provided. The use of computers to control and co-ordinate emergency response have evolved from using static computer [126]to primarily using mobile sensors.

Event Death Toll	Location and Date
China Flodds	China, 1931 2,000,000 - 4,000,000
Hurrican Katrina	Louisiana, 2005 1836
Indian Ocean Tsunami	South-East Asia, 2004 229,866
Rwandan Genocide	Rwanda, 1994 500,000
Darfur Conflict	Sudan, 2003 400,000
Mumbai Terrorist Attacks	Mumbai, 2008 134
Mt. Erebus Tragedy	Antarctica, 1979 257
9 11	New York, 2001 2974

FIGURE 6-8. **Disaster Statistics.** This table shows how the scale of an emergency or disaster is very different in different parts of the world owing to different population densities and nature of the emergency.

The EARS system was proposed [127]in order to display the radioactivity releases from the Diablo Canyon Power Plant. Multi-radio nodes (MRN) have been proposed [128] in order to deliver mission-critical information for decision support in battle command systems. In this case, the sensors are developed from COTS hardware and is basically a modified wireless router capable of interfacing with a radio modem. Coastal sanctuaries are prone to threats and support for emergencies in this domain have been proposed by integrating oceanographic observation with weather and pollution forecasts [129]. Robots have been augmented with autonomous navigation capabilities in order to enhance preparedness for chemical, biological and radiological/nuclear incidents at nuclear power plants [130]. Using the web for chemical emergency response [131] explored the possibility of allowing intelligent web services to be created to help the first responders in chemical hazards. Natural disasters such as debris flows have been monitored [132] with the help of real-time mobile multi-media communication systems. The utility of mobile phones has been studied for various scenarios including elderly life support [133],issuing early warnings using cell-broadcasting on GSM phones [134], enhancing positioning algorithms to locate activity amongst phone users during emergencies [135], 3 Using Cell Phones for Emergency Siutations The scale and severity of emergencies vary with the location in which they occur, the nature of the emergency itself and the population density, in case the emergency affects a set of people. Several examples come to mind from news reports including the Tsunami in the Indian Ocean, the Mumbai terrorist attacks, Hurricane Katrina and the Air New Zealand Flight 901 crash on Mt. Erebus. These examples typify different locations, population densities and communication patterns and requirements in the span of their occurrence. In most cases, the impact of a disaster or emergency is measured by the number of lives affected by it. Figure 6-8. presents a list of some of the deadliest disasters in history along with the metric used to measure them, human lives. Depending on the scale or type of the disaster, the requirements on the sensors used to aid in any part of the emergency continuum (Fig. 6-2), vary. For example, in order to gather information and data pertaining to an emergency, several methods can be used including profiling populations, monitoring weather forecasts, detecting suspicious activity either at borders of countries or other places where intelligence sources are concerned about trouble, etc. In order to prevent emergencies, systems in place would include early warning systems in order to evacuate people in preparation for large-scale disasters [131], sirens and of late technologies such as cell-broadcasting for emergency evacuation. In the case of deploying response units, several parameters are required in order to gauge the needs on the ground. For example, when Hurricane Katrina occurred, FEMA ordered 91 thousand tonnes of ice cubes [136] to cool food, medicine and sweltering victims, a week after the storm. This requirement is highly un-intuitive for any response team, unless they receive feedback from the affected people or other emergency management team members on the ground.

6.5 Conclusions

When considering emergencies, it is important to keep in mind that the metrics that determine the success or efficiency of a solution will vary depending on the scale and the place at which the emergency occurs. Since the nature of emergencies is so varied, it is important to accept that one size will not fit all. Our framework tries to incorporate various metrics, provide alternatives in the tools and solutions that we might adopt to solve particular cases and also keeps in perspective the underlying research problems that we might want to address, in the process. Cell phones are arguably the most ubiquitous sensors of human response or condition, even though their traditional role is not perceived that way. Given the density of hardware resources on a cell phone and the fact that the number of subscribers is nearly 60% of the planet makes a good case for using a tool that is widely deployes to proactively crowd-source civilians for relevant information in the time of an emergency.

7 Prototype

To investigate some of the ideas developed in the previous chapters, this chapter examines my prototype for Manikarnika. The elements I explore in this prototype are taking on the approach of providing a *framework*, which allows for easy adaptation across research questions, technologies, data and metrics. The bigger research questions that contribute to this framework come from the areas of operations research (specifically logistics) and emergency response (using a combination of position and reputation to evaluate better response mechanisms). In my prototype, I focus on emergency response but demonstrate that this is easily extended to OR. Technologies supporting my ideas that are part of the prototype include utilizing a few basic ideas to determine location and evaluating four different types of *schedulers* to understand how best to respond to a cross-domain emergency. I use a hybrid data-set, including caller data, cell-phone tower location data, raw performance measurements indicating RSSI values, visual components such as the map of the location being simulated in the prototype and actual data from real emergency statistics in the areas of interest. I present a unified platform for evaluating various metrics such as performance indicators including load on a tower, average call times etc., spatial expressions such as the time it takes the resource to cover a certain distance to get to an emergency event, etc. Section 7.1 discusses the overall architecture of the prototype, section 7.2 discusses my work on mapping user mobility over different time-lines and during different events. Section 7.3 details the work done on simulating emergency events, scheduling resources for those events and evaluating the effectiveness of the solutions. Section 7.4 talks about conclusions and future work.

7.1 Overall Architecture

In order to understand user-mobility and plot where the cell-phone tower resources were located, I initially used OpenSceneGraph along with static terrain maps of the South Island. The system to map user-mobility in my initial simple visualization of call loads and user movement is shown in Fig. 7-1.

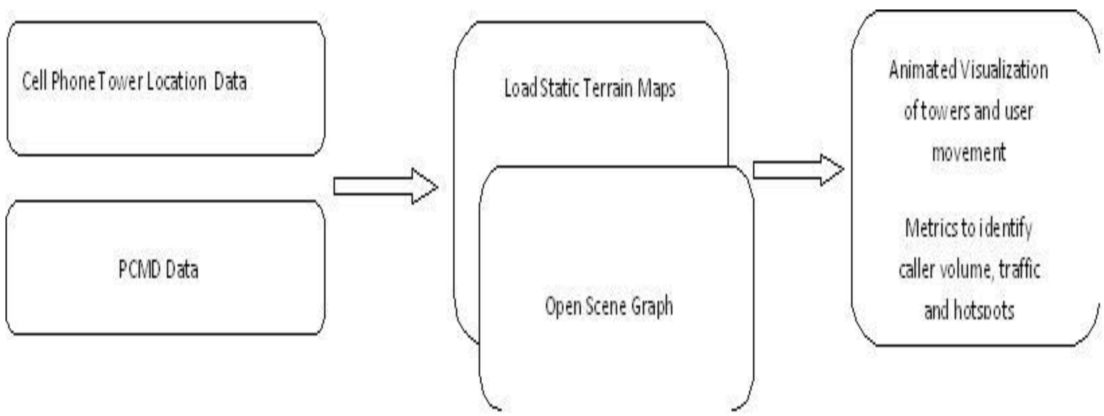


FIGURE 7-1. **Static Visualization - CAVIAR.** My Initial Prototype included a lot of static components such as terrain maps, only two sources of data (tower location and PCMD) and produced animated visualizations of user-mobility and tower-load.

I called this initial tool, CAVIAR, and it was very limited for a variety of reasons, including the infeasibility of adding new information in a layered fashion, defining boundaries of various administrative domains and encoding the notion of *time* in a more intuitive manner. These limitations lead us to modify the basic technologies I had assisting us, to and in the second version, for the front-end, I build my custom menus to simulate emergencies *in any part of the globe* wherein the NASA World Wind virtual globe is an integral part of my system. The NASA World Wind virtual globe is part of an open-source initiative where USGS satellite imagery is overlaid with several maps from ariel photography and other publicly available GIS data on 3D models of the earth and other planets. Several add-ons are available as part of this software including point layers, trial layers, line and polygon features, model features, place names, image layers and scripts.

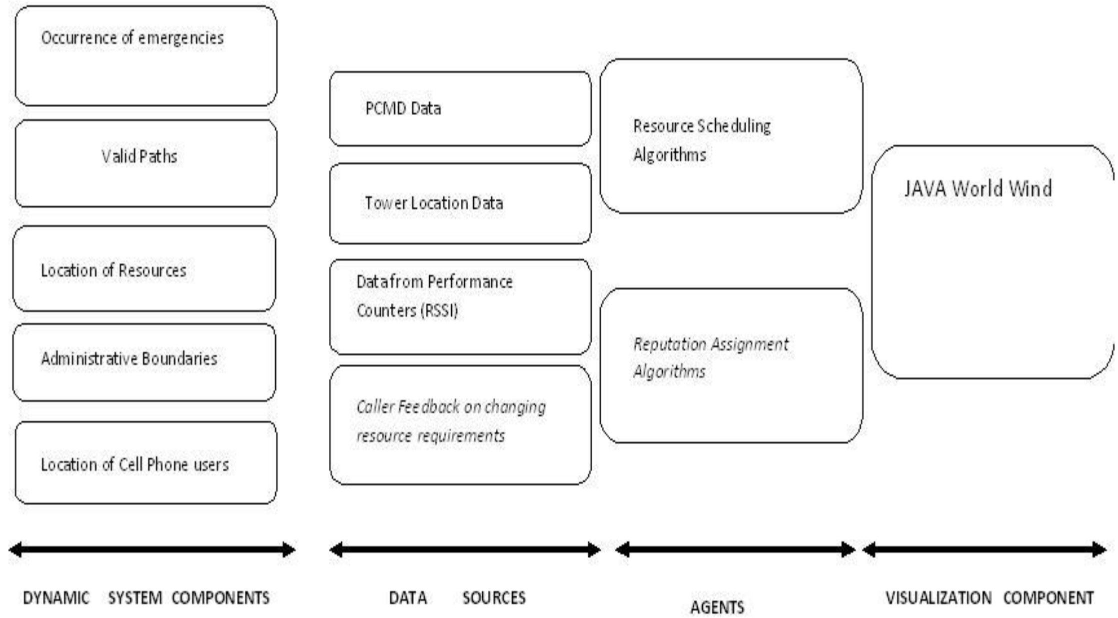


FIGURE 7-2. **Dynamic Modular Framework - MANIKARNIKA.** I introduce a mode modular design taking into account my framework for emergency response.

In my architecture, I have special configurations files to simulate various scenarios within the emergency management framework. All emergencies are not created equal and for example, an earthquake will lead to a clustering of emergency events along a fault line, vs. a bio-weapon attack will have a different distribution of events as compared to a tsunami, where the events will be clustered along the shore. Additionally, I am able to dynamically configure *invalid paths* such as resources being unable to traverse over water bodies or gorges or roads that are blocked as a consequence of the event. Since these configurations are *dynamic* I enable constant feedback (from say, mobile phone users that are being proactively crowd-sourced for information) to modify the parameters within which my resource scheduling and planning has to work. Fig. 7-2 shows the overall layout of my new framework.

7.2 User-Mobility

Utilizing my first attempt at the design, I come up with a simple visualization of the cell phone towers (shown in Fig. 7-3), tower-statistics (shown in Fig. 7-4) and user-mobility (shown in Fig. 7-5).

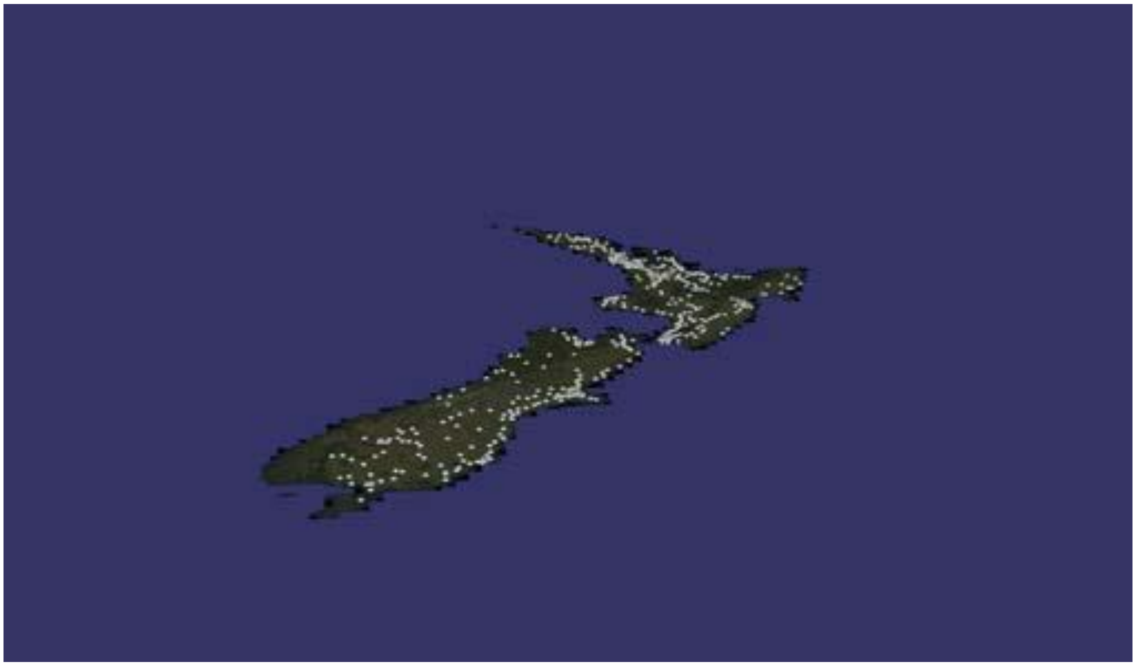


FIGURE 7-3. **Cell Phone Towers.** I show the location of cell-phone towers in this figure, from my simple visualization tool, CAVIAR.

Cell phone tower locations are contained in their own data set outside of PCMD. I am provided with unique identifiers for each of the cell phone towers and a (latitude, longitude) for each tower. I convert these into (Easting, Northing) for each tower. Easting and Northing are terms to describe what are essentially geographic Cartesian pairs for any point. Taking a base map of New Zealand, I plot these points on that map in order to depict the location of the cell-phone towers. It is easily observed that the North Island has a lot more towers than the South Island, owing to a larger subscriber base. Added to this, the grouping of towers is also indicative of the underlying terrain, for example, in the South Island, there are not too many towers in the regions housing the Southern Alps. This is not obvious from this visualization, however as there are no terrain depictions of the region.

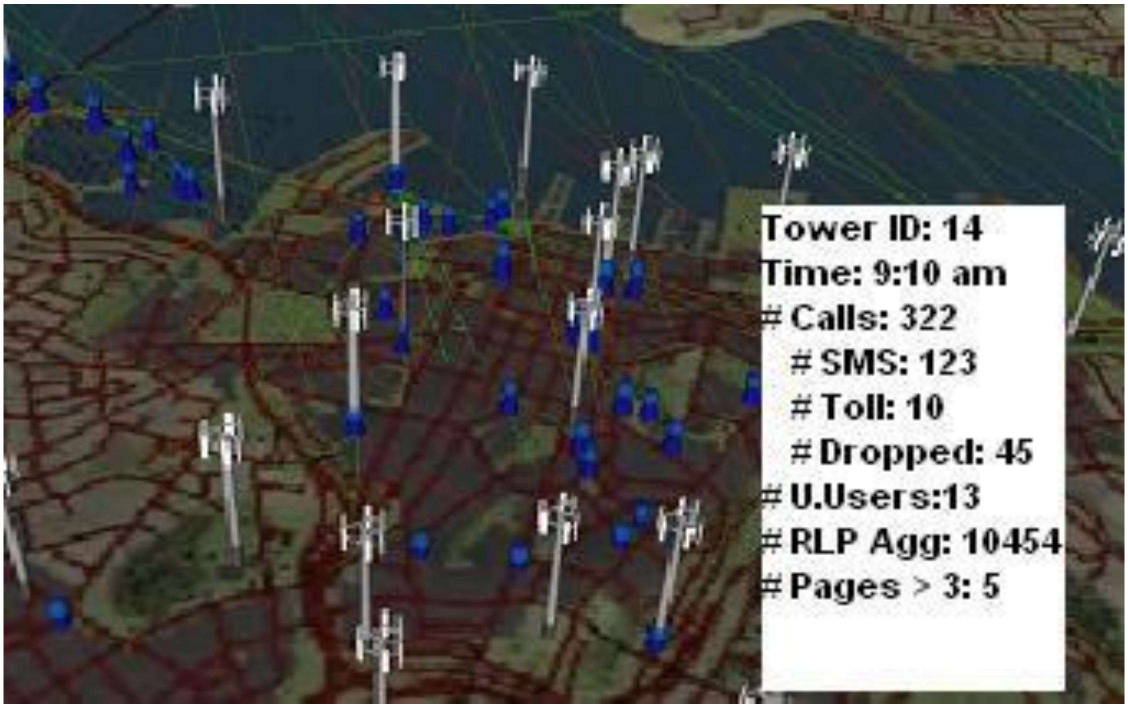


FIGURE 7-4. **Tower Statistics.** Cell Phone towers are the white structures, each of the users around the tower being depicted in blue. On rolling over a tower, simple statistics about the tower such as the Tower ID, the Time at which the query is being processed and other data related to the tower is displayed.

In this tool, I depict phone subscribers with blue icons and enable basic zoom and roll-over functions and Fig. 7-4 shows the zoomed-in view where the user has rolled over a tower and the basic tower statistics are displayed to the user. In my particular example, the statistics I show the user (in this case, an emergency responder or a network capacity planner) such variables as the Unique ID of the tower, the time-stamp which the data-shown pertains to, the number of calls made, the different kinds of calls made (SMS, Toll calls, dropped calls), the number of unique users whose calls originated on the tower, the number of paging requests seen (which could provide insight into security violations on a tower) etc. The proximty of the subscirber to the tower is calculated using the call duration and the call start time and the ending cell tower. This is a static snapshot at 9:10am and does not take into account how the user is moving. Further, the orientation of the user is approximated based on movement and if the user is static, I can only estimate the distance from the center of the cell-tower.

FIGURE 7-5. **User-Mobility.** Cell Phone users are shown moving at different rates (green being slowest, red being fastest and yellow and orange representing the range in between). This visualization is not effective in encoding time.

Fig. 7-5 shows user-mobility between cell-phone towers. From section 3, I know that most users do not move very much, while making their phone calls. 54% of the users keep their movement to within two cell towers (Fig. 3-4). In this example, I have filtered out the users that are moving between more than 2 cell-towers in the course of their conversation (around 20% of the total trace). By further filtering the data to only include a few *minutes* worth of calls, I am able to depict the users moving, along a certain trajectory (shown by green, orange and yellow lines) on the map. How quickly a user is moving is calculated based on the starting and ending tower for a particular call, the length of the call and the location of the starting and ending towers. I depict the *shortest path* that a user can follow, during the course of the call. This is for purposes of visualization only as the user might well take a circuitous route. Added to this, my depiction of the user's movement does not address some of the infeasible paths such as traversing what are obviously water-bodies (shown in the top left-hand corner where I see users freely zooming over Lake Taupo while placing a phone call). For the initial studies on user-mobility, this depiction is sufficient for estimating call traffic and call volume as I will observe in coming examples.

This visualization was an initial step towards characterizing call traffic and user-movement over the course of a day. In this part of the prototype, I successfully compared how call traffic differs between a normal day in the week versus a day on which an All-Blacks Rugby game was played in Christchurch. The data-sets used for the examples shown in Fig. 7-6 and Fig. 7-7, which contrast these two days are shown in Table 7-1. The data was collected in hour-long blocks for my visualization.

Trace	**Date**	**Event**	**Sample Size**
One day dataset (Christchurch)	22nd January 2008	Normal Day	1667914
One day dataset (Christchurch)	21st June 2008	All Blacks Rugby Game	1272184

TABLE 7-1. **Data-Sets used to show call-traffic.** I compare traffic over 24 hours, on two days, a normal day and the day and All-Blacks Rugby game is played in Christchurch.

Fig. 7-6 shows the call traffic within Christchurch, from the first data-set shown in Table 1. The data is broken up into hour-long samples, for one whole day, January 22nd, 2008. This is a normal weekday in the city with no special event occurring anywhere to create anomalous traffic patterns. Each tile represents an hour, starting at 8am with 12 tiles being displayed for 12 hours of the day from 8am to 7pm. The visualization is straightforward in that the density of calls is depicted by the size and the brightness of the circle depicting that statistic, which is placed over that particular cell-tower, in the city.

FIGURE 7-6. **Normal Day Call Traffic.** Call traffic is depicted within Christchurch on January 22^{nd}, 2008, a normal day, over 12 hours.

As can be seen from Fig. 7-6, at 8am, the city wakes up and from 9am-5pm,the city is most active, which pertains to the work-hours of people residing in the city. Also seen is the concentration of calls in the cell towers closers to the city centre, which is starting to brighten (depicting greater number of calls) at around 9am, which is the second tile in the topmost-row of tiles in Fig. 9. At around lunch hour 12pm-1pm, I see that the traffic starts to move to locations outside the city, possibly commuting to get lunch at spots outside the city-centre. This is depicted in the third tile in the second row. After about 5pm 5:30pm the calls drop off as night-time approaches.

FIGURE 7-7. **Game Day Call Traffic.** Call traffic is depicted within Christchurch on June 21st, 2008, when an All-Blacks Rugby game was played, over 12 hours.

Fig. 7-7 shows the data collected for calls made on June 21st, 2008, corresponding to the date of an All Blacks game vs. England. The difference in call traffic here depicts how the city behaves during a holiday or during a special occasion. At 9am, the city is awake but not buzzing, as on a regular weekday, which was January 22nd, 2008 (a Tuesday). Furthermore, three is no migration of call traffic because most users have decided where they are going to be, at the time of the game, so the intensity stays static between tiles 3-8, depicting the hours of 10am-3pm. At the time of the game being played, 7:35pm was the telecast time, depicted in Fig. 7-7, in tile 1 in row 1 and for a few hours preceding that, depicted in Fig. 7-7, tiles 1-3 in row 4 (showing 5pm-7pm), the call traffic dies out as people are more focused on the game. This, in contrast to the same tiles 1-3 in row 4 (showing 5pm-7pm) in Fig. 7-6, is much less call traffic. In Fig. 7-7, between the hours of 5pm-7pm on a weekday, the city was still pretty awake and call traffic depicted that. I have co-related this observation to other All-Blacks games played in Wellington and Auckland but do not include these for length restrictions.

7.3 Reverse 111

In this section, I describe my more modular design that incorporates the *framework* while allowing for diverse data-sets and metrics to be measured, using the same tool. The objectives of building this tool are to evaluate a number of ideas and propose a dynamic design-framework to achieve this. By tying in the concepts of user-mobility and resource schedulers, I evaluate the utility and relevance of considering a proposal such as Reverse 111 and provide a visual tool to assess my proposal. Fig. 7-8 shows the opening screen of my prototype.

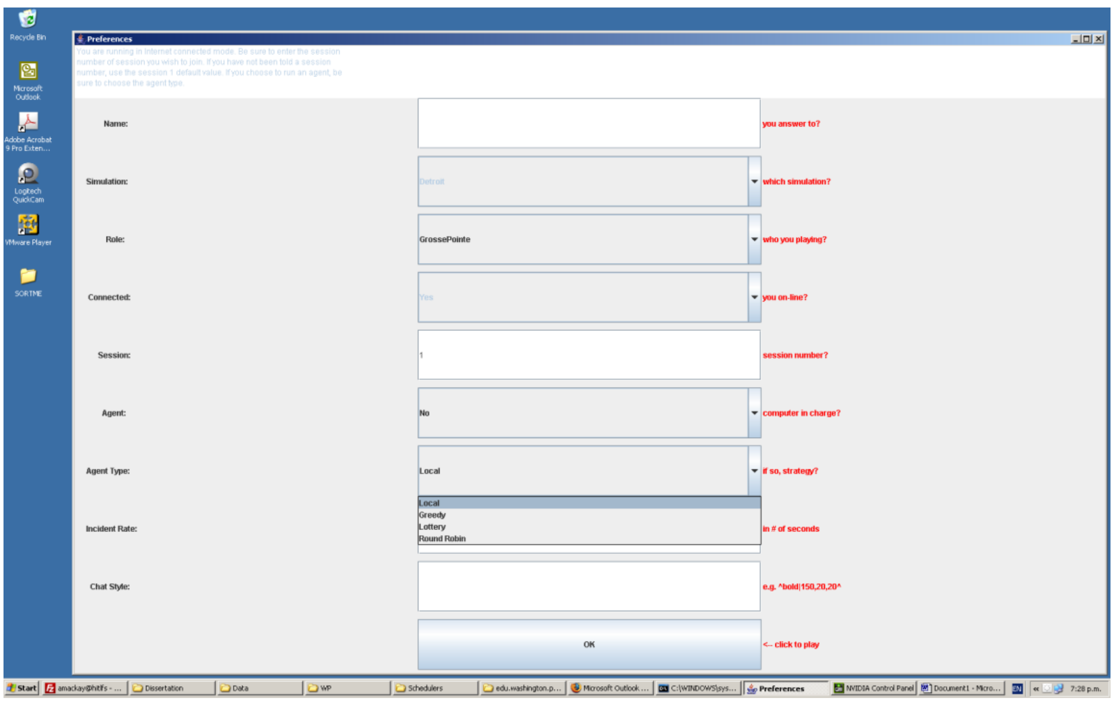


FIGURE 7-8. **Opening Screen.** Users of this framework are able to configure a lot of system parameters on the fly including roles assumed, playing with other online players, recording the history of their session, playing as human players or invoking scheduling agents that evaluate the effectiveness of various scheduling ideas, controlling the incident rate, etc.

The dynamically configurable aspects of the prototype are seen in this figure to include the region which the user wants to evaluate their proposal in, which could include any geographic region in the world. For purposes of example (where emergency situations can assume two different scales, I limit my discussions to New Zealand and the city of Seattle in the state of Washington, in the United States). The user can choose to run the simulation as an *oracle player* who controls all administrative domains (which is useful for baseline comparisons) or as a particular administrative domain (such as the mayor of Christchurch), in order to understand which policies work best, given their resource constraints. This tool also allows multiple-players connected online to play simultaneously, but that is not central to my use. The user's actions and findings can be recorded on a per-session basis, creating a powerful tool for analysing historical results and comparing these with moving the parameters of the simulation around. Different types of resource-schedulers (or agents) can be evaluated in this tool and in case the user does not want any agent, wants to schedule resources themselves (which was the original design of a visual analytics end-use of this project []), then they can play as themselves, human users. Users are also able to choose the rate of occurrence of emergencies (once more, the emergencies will have a notion of *scale* associated with them and an emergency in India affects a lot more people and occurs quicker and more frequently if wide-spread, as opposed to an emergency in New Zealand) to keep my results realistic, when evaluating various agents.

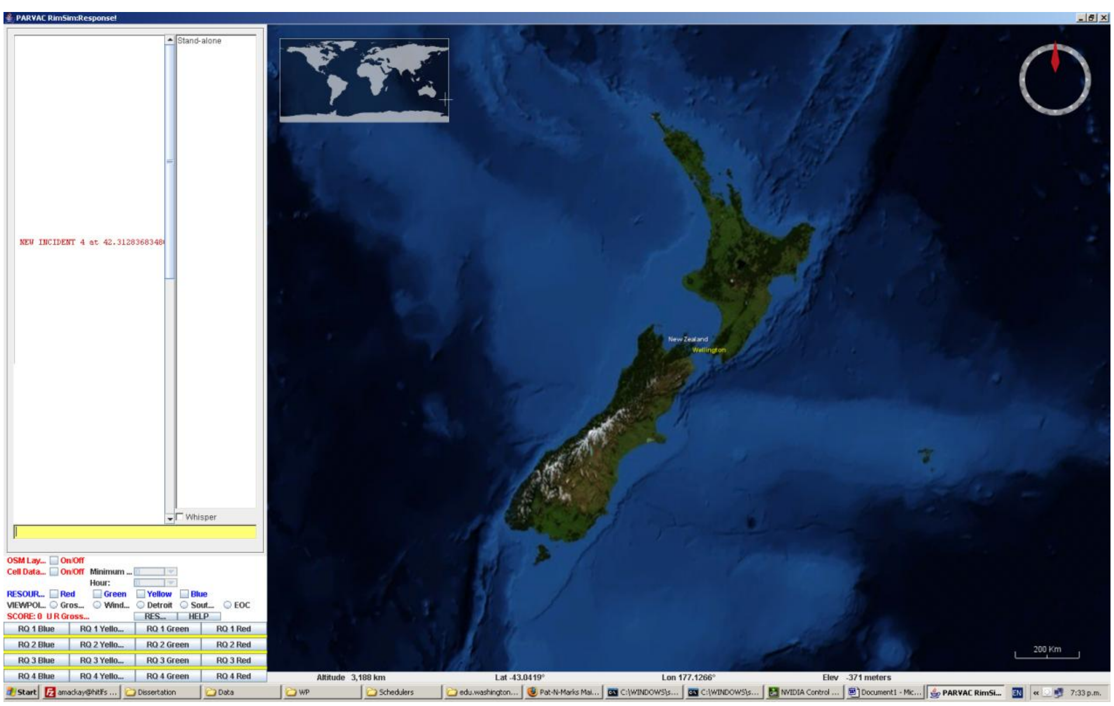


FIGURE 7-9. **Layered Approach.** This shows the first level view of my tool wherein the panel on the left-hand side allows the user to see the data and information that they need, selectively.

Fig. 7-9 shows the presentation of the top-level screen as the user enters the starting phase of the simulation. I take a layered approach by presenting the data and information, as the user sees fit, rather than all in one piece, as I did with the user-mobility aspect of the prototype. The user is notified with messages on the top-left panel and allowed to choose which of resources, incidents and *cell phone data* that they wish to see appearing on the screen. They may choose to see one or the other or some combination of this information, with the progress of time. I also add appropriate labels to indicate the actual location that the user is viewing in the OSM layer, as a reference point. This can be extended to mark administrative domains as well. A layered approach is a direct consequence of my modular framework.

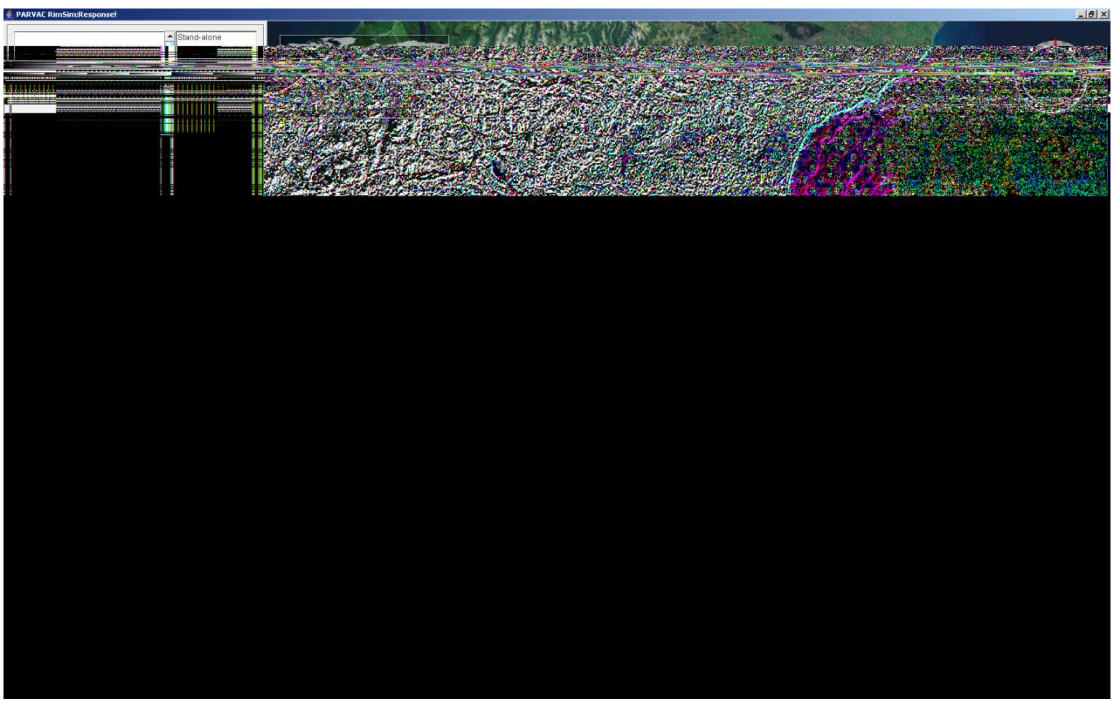


FIGURE 7-10. **Zoom In.** This shows the first-level zoom-in function, which shows the exact latitude and longitude on the bottom panel at a resolution of 20km, rather than 200km from Fig. 7-9.

Fig. 7-10 shows the first level zoom in, which opposed to my static visualization tool, CAVIAR, shows a lot more information, including the scale (20km) and the Latitude and Longitude where the zoom is occurring. This is important for user-orientation. Furthermore, this visualization environment takes into account the terrain, much better than using static maps, thereby making it easier for the user to identify *valid paths* and administrative domains. In my earlier examples, these were difficult to identify, specifically elevation, but that is shown to a greater detail in this environment.

Fig. 7-11 shows the setting of *administrative domains* demarcated by the magenta lines. In this case, I have Christchurch, Rangiora, Oxford and Ashburton as the four different administrative domains that must interact and co-ordinate resources at the time of an emergency. This figure also shows the introduction of emergency events at an interval of 4 seconds each, indicated by the rhombus-shaped objects, with the four colours standing for one kind of static resource with numbers inside them indicating the resource-requirement. For example, if a rhombus appears with 3-yellow,0-red,1green and 1-blue, that is the resource requirement for that emergency, at the time it is occurring. This is shown in Fig. 7-11 upon zooming in on the resource. This requirement can *dynamically change* if a co-located cell-phone user, indicates as much back to the system. The resource requirements are important to assess the scheduling policy's action.

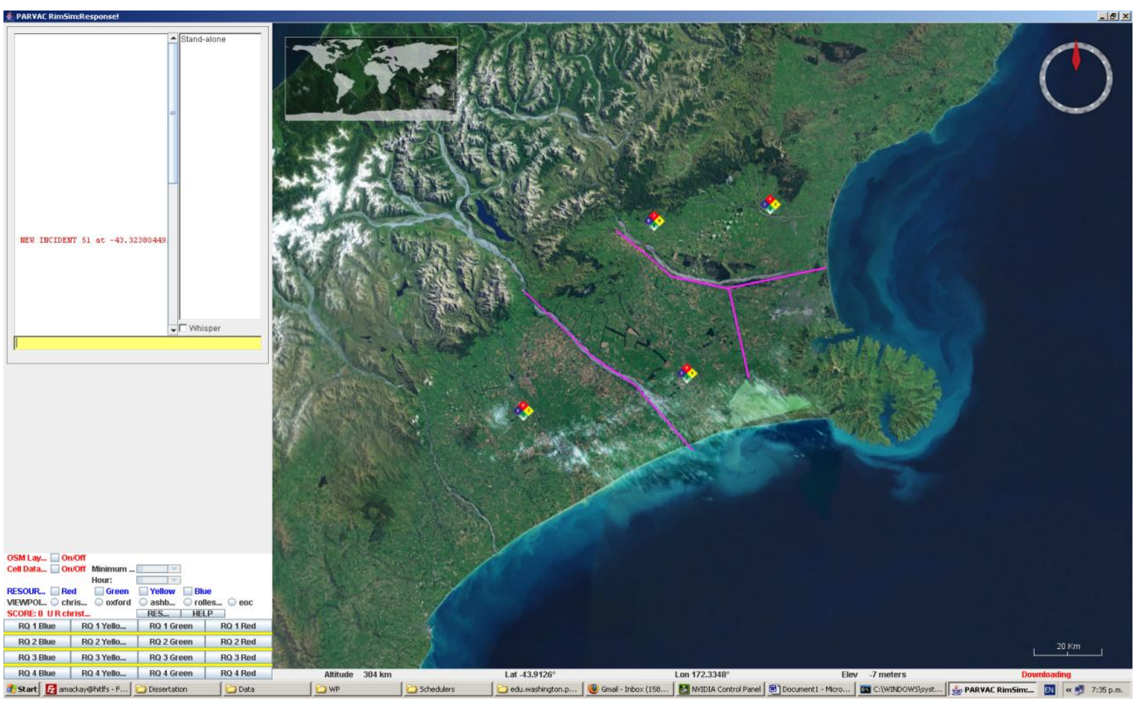


FIGURE 7-11. **Administrative Domains.** Four administrative domains, Christchurch, Oxford, Rangiora and Ashburton must act in tandem for this simulation situation.

Events occur at pre-determined locations or according to various pre-simulated scenarios. Fig. 7-12 shows the occurrence of three different emergencies within Seattle, including an earthquake, a bio-weapon attack and a tsunami, which have different distributions and groupings of events along the geographic spread of the region.

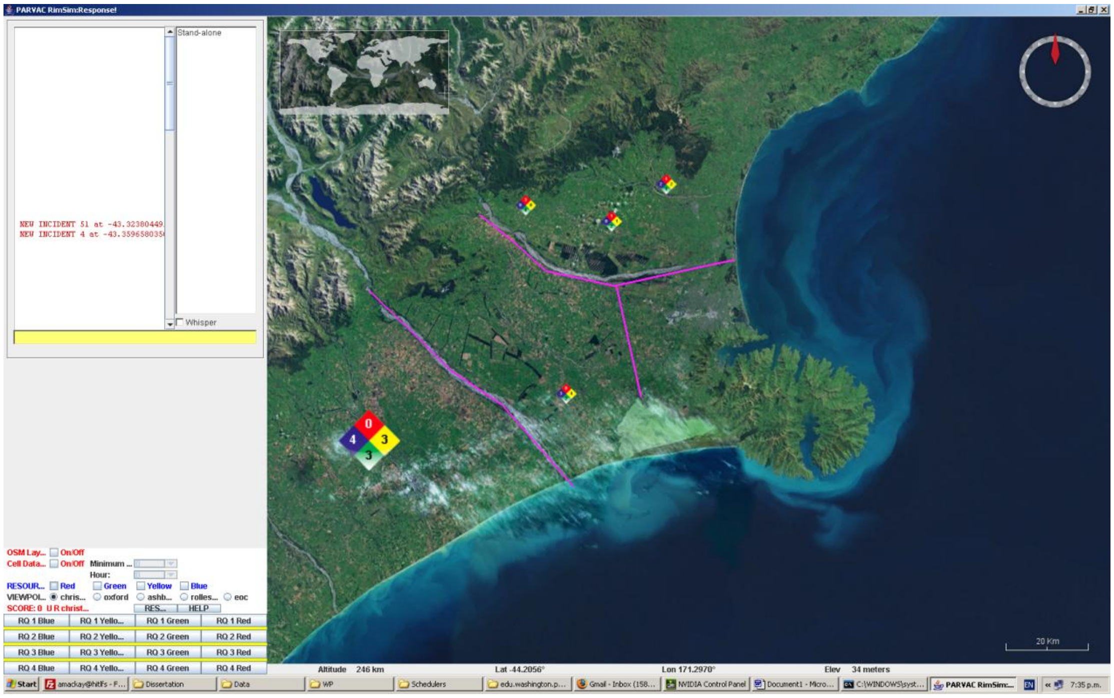


FIGURE 7-12. **Event Zoom.** Upon zooming in on an event, its resource requirements are shown to the user. In this case, the event required 4-blue resources, 3 yellow, 3 green and zero red resources.

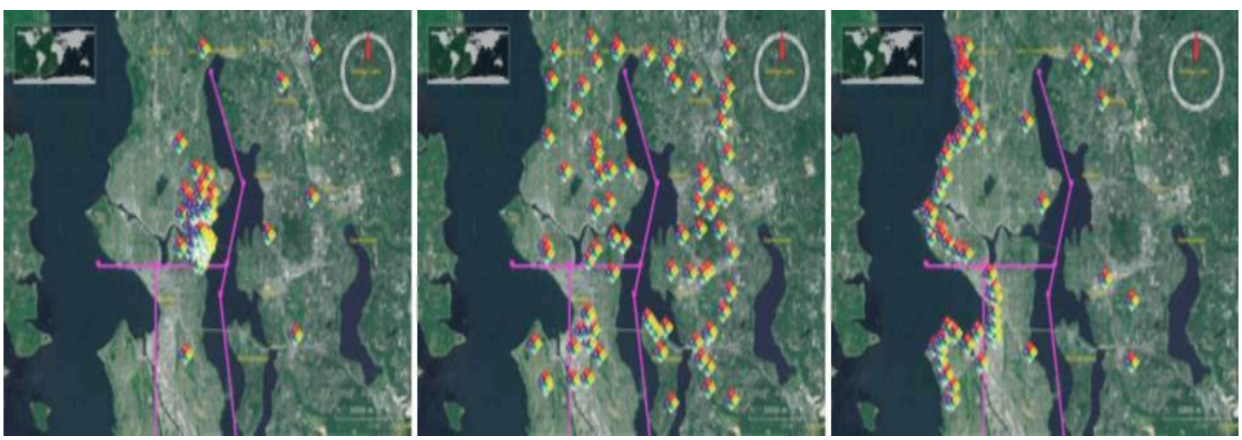


FIGURE 7-12a. **Different Event Distributions.** Three different types of events including a bio-weapon attack, an earthquake and a tsunami show different occurrence and grouping of incidents in the Seattle area.

While I take into account actual numbers of fire-trucks and other resources that have been reported on public websites accounting for the State Department's response to such activities, I combine this with an intuitive approach to where and how events within an emergency scenario might occur and my visual representation confirms that my placement of events is accurate.

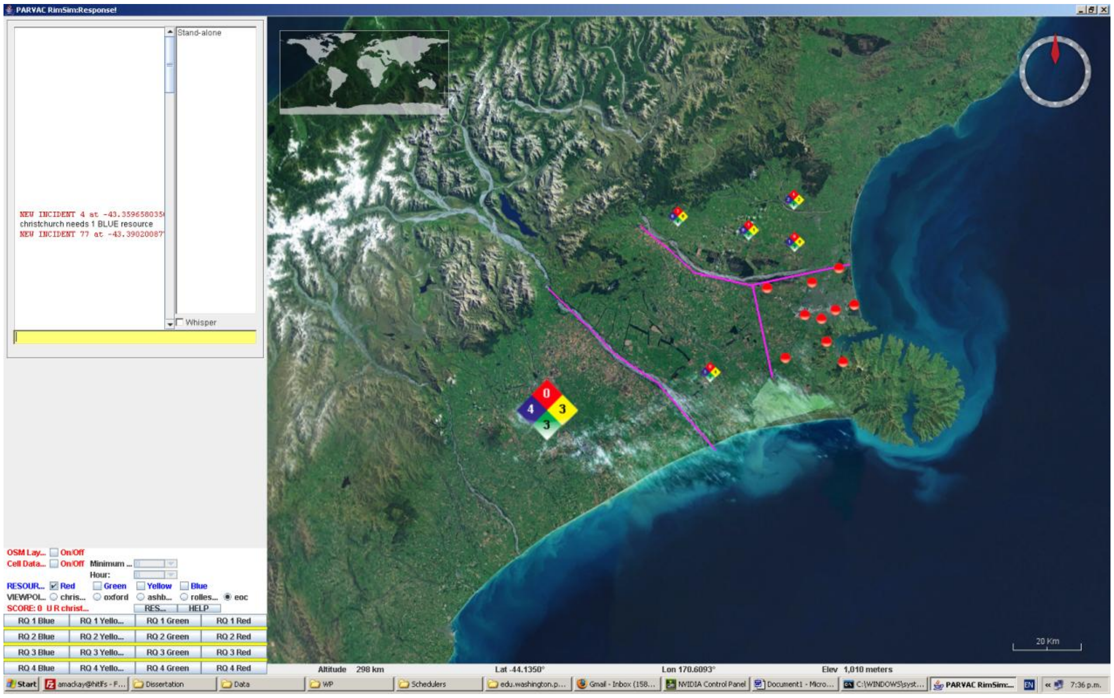


FIGURE 7-13. **Resource Layers.** The user can select to see the actual physical location of resources, in this case they have chosen to see the location of all the red resources. These resources are applied to Emergency management in this example but can be easily extended to logistics and visualizing Supply chains.

Fig. 7-13 shows the location of resources of a certain kind, when the user selects that resource should be shown on screen. The location of resources also gives the user a visual clue on the performance of my schedulers, when they try to pick the resource that is suited, according to their policy, to send to a particular event. Fig. 7-14 shows multiple resources (of various colours on the screen, as selected by the user. In this view, the user has also chosen to play as the administrative domain of Christchurch, therefore placing a greater emphasis on that region alone.

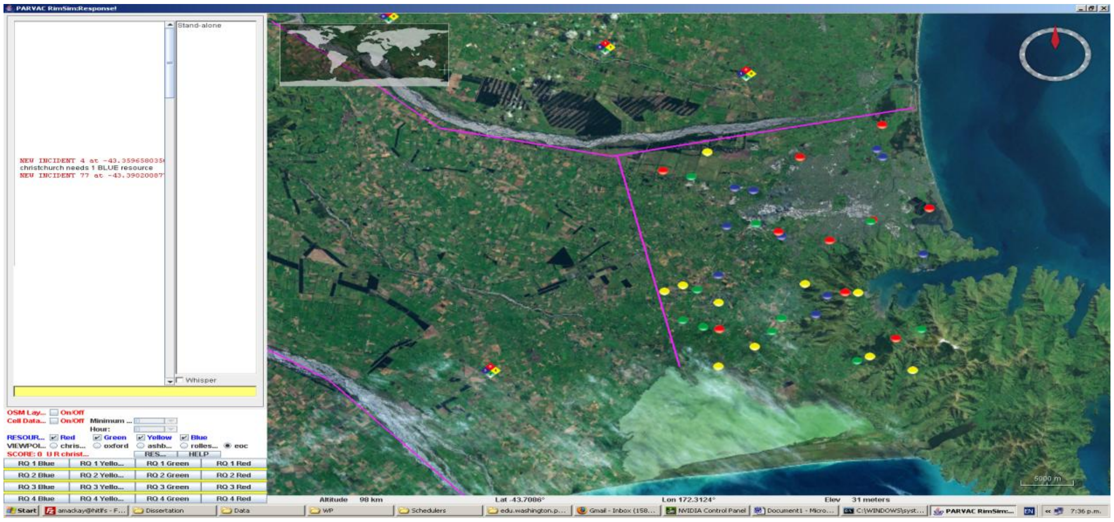


FIGURE 7-14. **Resource Layers for Christchurch.** The user has selected to view all resources and play as the administrative role of Christchurch, rather than all four domains indicated in earlier figures.

7.3.1 Path Selection

In this section, I discuss how paths are selected in order to get a resource to an event that requires it.

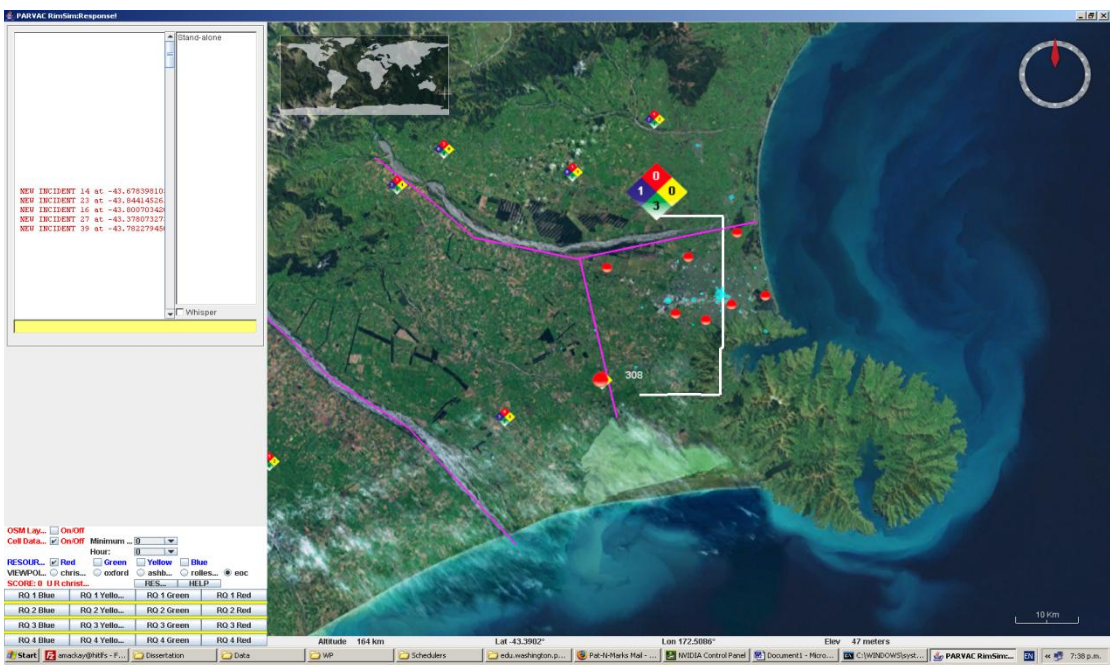


FIGURE 7-15. **Cell-Phone data Christchurch.** The user has selected to view all resources and events, and the cell phone users within Christchurch, indicated by the light blue circles. In this example, an event is being handled, with the white line indicating that a resource has been scheduled for the event that is highlighted at the end of the white line.

Fig. 7-15 shows the scheduling of a resource for an event that is highlighted at the end of the *path* for that resource and the cell-phone users in the Christchurch area, shown in light blue. Cell phone users are considered *soft resources* that provide *information* as detailed in Chapter 5. In my prototype, the path selection process goes through the steps described in the flowchart shown in Fig. 7-16.

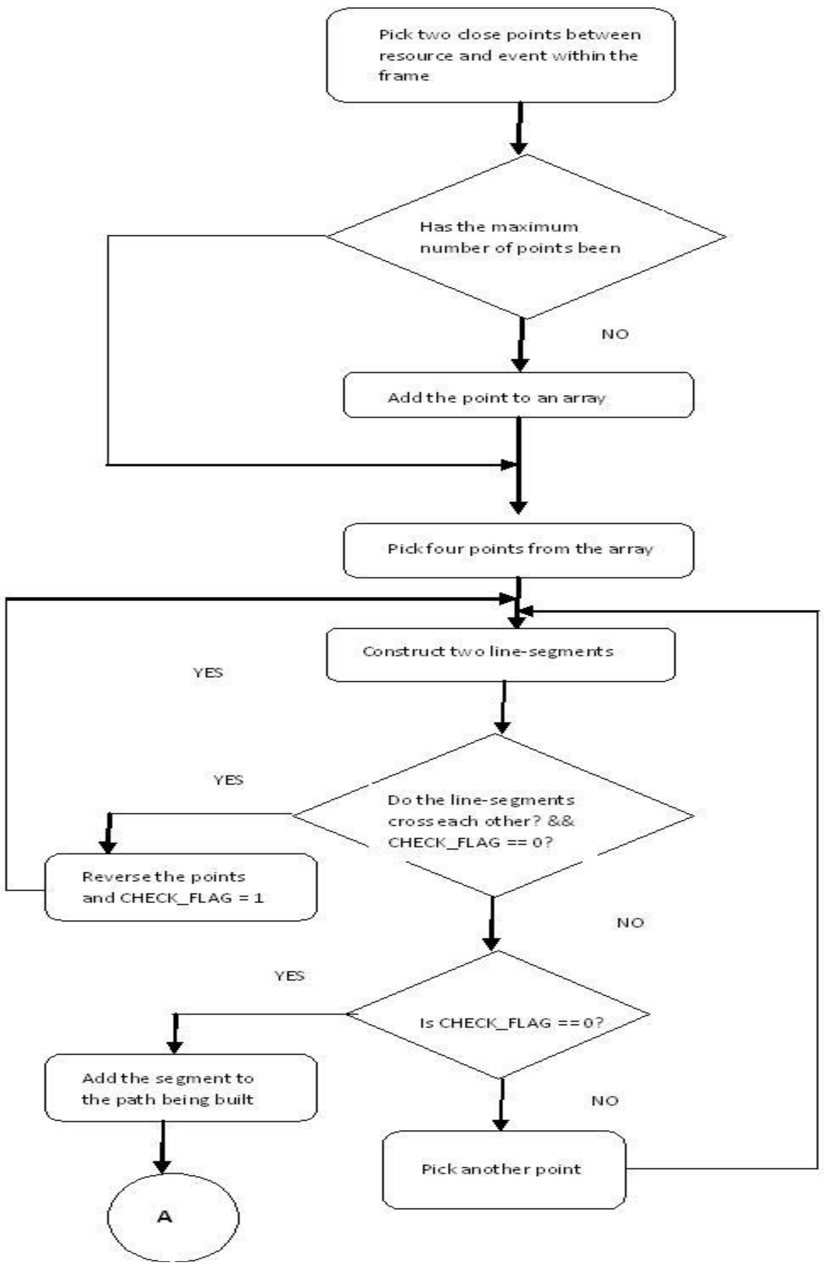


FIGURE 7-16. **Building a Path.** The system constructs a path in a piece-wise fashion for each valid FRAME, which is only within the region that can be traversed without intervening geographic hindrances like water-bodies, etc.

I draw the reader's attention to Fig. 5-9 wherein the user sets up the specific regions or *frames* that are valid for the resource to traverse. My system constructs a piece-wise linear path, for each frame of valid points, located within a sub-region that can be navigated without any obstacles in the resource's way such as water-bodies or gorges (in the case of New Zealand's South Island). Fig. 7-16 shows the construction of the path, while ensuring that none of the segments cross each other, which would make for a circuitous route for the resource, as it reaches the event of interest.

FIGURE 7-17 **Traversing a Path.** Once the path is confirmed, the resource traverses the path to the event of interest, as shown in this figure.

Fig. 7-17 shows a resource travelling towards an event of interest. The path is now shown in yellow as it has been fully constructed and confirmed, whereas the path in Fig. 7-15 (in white) was still under construction at the time the screen-shot was taken. Fig. 7-18 shows a co-located user, that is directly on the path being traversed, that can provide some insight into the situation towards which the resource is travelling. The user on this path has *reputation information* associated with them, besides actual physical location. For example, the user is associated with tower #27, which has seen a total call volume of about 1000 calls, 276 from distinct callers. This tower is heavily loaded at the time of the emergency (I am tracking an hour of call-time in this example and filter only users that have not moved across cell-towers, indicating static users and not perpetrators that are fleeing) and it is likely that this user is reputable as the average call time of calls on this tower are about 22seconds, indicating that most calls are being placed to check to see if the called party is safe.

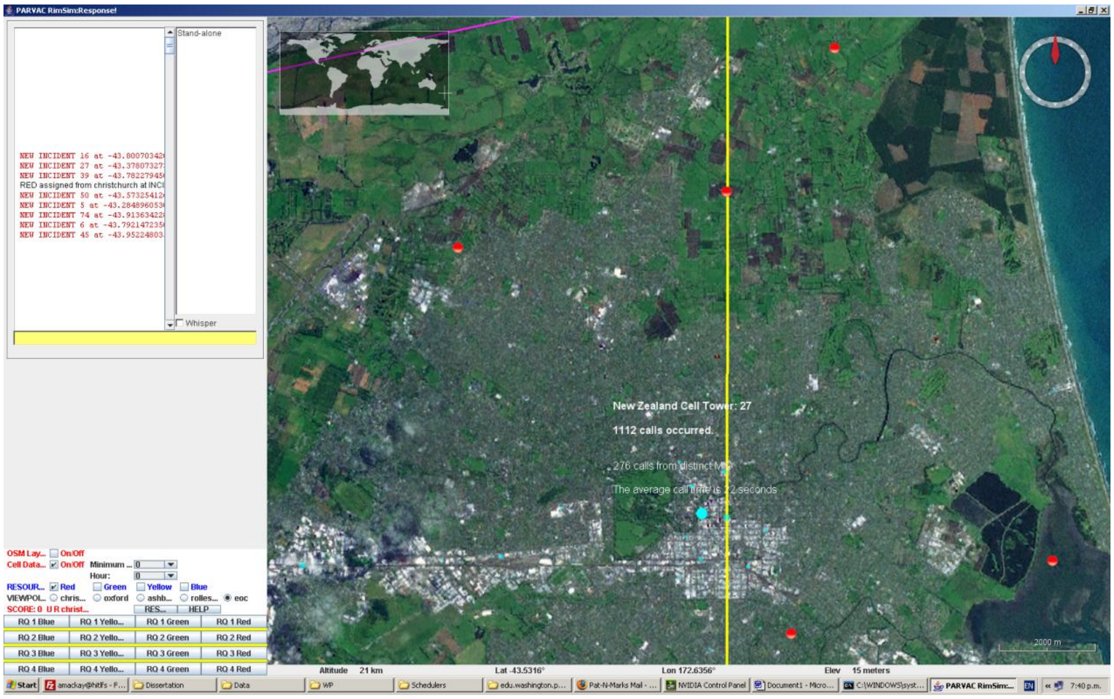


FIGURE 7-18 **Location + Reputation = Information.** This figure shows obtaining information from a reputed user, co-located with the emergency event and modifying the resource requirements to take that users-input into consideration.

7.3.2 Scheduling Agents

In this section, I perform a different evaluation of *four* scheduling agents, which was joint work with the University of Washington. I pick the three scenarios introduced in Fig. 7-12, for this evaluation, in the greater Seattle region. The scale of emergencies in this region is different, as is my evaluation method, when compared to the result presented in Chapter 5.

In order to simulate multiple first responders within a response scenario, I added an agent class that simulates four different schedulers, greedy, lottery, round-robin and first-fit. In order to encode realistic scheduler behaviour, I used a combination of resource-sharing policies, and resource and event selection strategies. These heuristics are easily extensible to encode newer policies and many more scheduler types, either distributed or otherwise.

The emergency response behaviour of sub-classed agent types are varied by tweaking three parameters that affect agent behaviour: (1) their policy on sharing resources with neighbouring administrative domains (for example, does the mayor of Christchurch share with the mayor of Rangiora, etc.), (2) their strategy for determining which active unresolved incident should be attended to next (i.e does the scheduler choose the first event which appears or perform calculations to see which event can be responded to with the closest resources), and (3) their strategy for determining which available resource to assign the targeted incident (i.e, if an event requires 3 red resources, do I get all three to the event in one go or do I phase out the resource delivery).

Each agent takes a turn in effecting the dynamic visualization until the scheduled simulation time ends or all incidents have been handled and no more are scheduled to occur. A turn is calculated as follows:

1. An agent figures out its resource surplus (how many resources it has free, minus the number of unmet resource demands from incidents in its jurisdiction). I call this number **R** (note: R is occasionally negative).
2. I then interrogate the agent's R-level willingness to help other administrative domains. I call this number **HiWat**. If R > HiWat, the agent looks at the requests for help that have been made by other players (human or agent). If there is an open request and the R > HiWat, the agent chooses a resource according to its Resource Selection Strategy (RSS) and send it to the incident specified in the request. This ends a turn.
3. If instead R <= HiWat, I interrogate the agent's R-level willingness to ask for help from other administrative domains. I call this number **LoWat**. If R < LoWat, the agent chooses the last incident on my priority list (if there is one; priority as determined by my ISS) and broadcasts a request for help on that incident. This ends a turn.
4. If instead R >= LoWat, the agent chooses an incident according to its Incident Selection Strategy (ISS) subject to its Resource Sharing Policy (RSP). This RSP provides a primary

ordering of the incidents according to jurisdiction, the ISS provides a secondary ordering in case of ties (and usually I find there are many ties). If there are no incidents to choose, this ends a turn.

5. The agent chooses a resource to send to the incident it has chosen by following its embedded heuristics, which in this case are according to an RSS. The agent sends the resource to the incident it chose in step 4. This ends a turn.

For the study I document in this thesis, the agent behavior-defining parameters I chose to test include:

- **HiWat** - an integer representing willingness to give help when asked.
- **LoWat** - an integer representing willingness to ask for help
- **Resource Sharing Policy** (RSP) - one of 5 alternatives representing willingness to volunteer help.
- **Incident Selection Strategy** (ISS) - one of 4 alternatives representing heuristic for choosing the next incident to apply resources.
- **Resource Selection Strategy** (RSS) - one of 2 alternatives representing the heuristic for choosing the next resource to assign.

7.3.2.1 Resource-Sharing Policy

An agent's RSS policy describes under what conditions it will voluntarily respond to incidents outside its given geographic region. Five policies are defined and implemented in agent sub-classes. I refer to the policies by name in order to assist reading comprehension when describing my experiments in a later section:

- Sociopath - Never volunteer aid to another region.
- Selfish - Prefers incidents in its own region, but will volunteer aid to another region if there are no active incidents in its own region.
- Equalitarian - Does not take geographic region into account when determining which incident to handle next.

- Selfless - Prefers to volunteer for incidents in another region, but will handle incidents in its own region if there are no outside incidents to handle.
- Altruist - Never handles its own incidents, but will always volunteer for incidents outside its region.

7.3.2.6 Incident Selection Strategy

Within the broader resource-sharing policy, there is still the question of which incident to handle first, since there might be many active incidents within a single geographic region at any given time. Which incident is selected in the end depends first on policy and then on incident selection strategy. I implemented four representative incident selection strategies for my experiments:

- First Fit - Chooses the incident with the lowest incident number regardless of other considerations. Computationally, this is far the simplest of the strategies.
- Round Robin - Chooses the incident that has been active the longest.
- Lottery - Gives each incident a number of tickets equal to the total number of resources it requires, and chooses a ticket at random. The incident holding the winning ticket is selected.
- Greedy - Considers the resources that would have to be applied to each incident, and chooses the incident that could be handled most quickly (that is, on the basis of the furthest required resource).

7.3.2.3 Resource Selection Strategy

Once an incident has been identified, an agent must choose resources to assign to that incident in order to assist in incident resolution. There are likewise many possible strategies for choosing between resources to assign. For the demonstration purposes of this thesis, I encoded two resource selection strategies:

- First Fit - Chooses the free resource with the lowest resource number.
- Closest - Chooses the free resources closest to the incident.

7.3.2.4 Agent Types

The agent types that I use to demonstrate my software use are a combination of the elements above, particularly the two selection strategies, since the policy can be supplied as a parameter for every agent type:

- LocalAgent: First Fit incident selection, First Fit resource selection.
- RoundRobinAgent: Round Robin incident, Closest resource.
- LotteryAgent: Lottery incident, Closest resource.
- GreedyAgent: Greedy incident, Closest resource.

7.3.2.5 Simulation Runs

Each of the scenarios is divided into the same four geographic areas to represent administrative domains for the emergency response agents. I varied all possible 160 agent characteristic mixes across all four geographic regions to create 640 runs for each scenario. I watched each of the 640 simulations take place for each scenario and noticed many interesting patterns of resource allocations including obvious inefficient motions for inferior characteristic sets. The response activity across the two major bridges was especially interesting to watch as were the clustering of movements at times between administrative domains. Figure 7-19 looks at the time to completion for each simulation across all three scenarios. Scenario completion times include the area from the top of each scenario area down to 0 on the x-axis. In all but twelve of the 640 runs, the tsunami scenario took less time to mitigate than the others (in twelve runs the completion time was identical to the earthquake scenario). The earthquake scenario took less time than the bioweapon scenario in all cases (but was much closer in the most efficient cases). The distribution of response times gives a sense of how significant agent behavior mix was to the completion time of the scenario. The more isolated the community event to one jurisdiction, the more impact agent behavior makes on completion time.

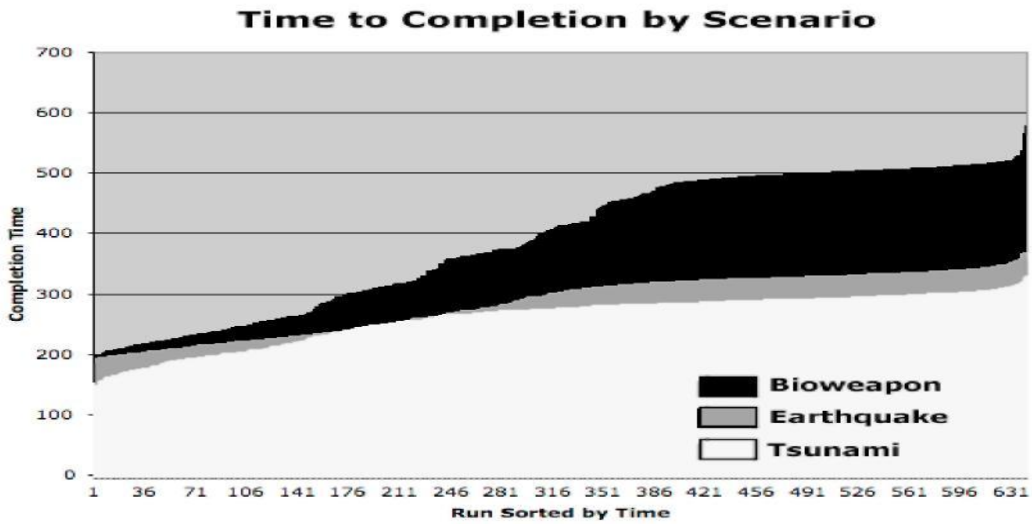


FIGURE 7-19 **Seattle results.** This figure shows my running various agents within the greater Seattle region. I do not take into account co-located callers in this example.

7.4 Conclusions

In this section, I present a multi-tiered prototype, which incorporates cell-phone information as one layer of the mapping tool that is part of the prototype. Detailed simulation of emergencies, resources available by region and various scheduling policies to assign resources to emergencies are presented in the prototype for thorough evaluation. By communicating between the location of occurrence of simulated emergencies and co-located civilians, I try to motivate the Manikarnika framework in general and the Reverse-111 idea in specific, to utilize proactive crowd-sourcing for Emergency Response. As opposed to proposal that use cell-broadcasting to notify subscribers *en masse* proactive crowd-sourcing is very specific in its recognition of who in the crowd the system contacts, in order to procure reliable information from, regarding emergencies. In future work, I would seek to explore estimating the *steady state* within an emergency, in order to address the subjective nature of these results. I also envision evaluating hybrid schedulers, whose policies can change depending on where in the time-continuum the response is being applied.

8 Conclusions

My thesis is that the use of *location* is central to enabling solutions that address the issues of scale, in providing meaningful solutions in both commercial and non-commercial applications, on the most ubiquitous *sensor* of them all, the cell phone. When looking for clues to establish how and when the location metric is useful, I performed a series of performance analyses in my trace-driven environment, which was very different from previous simulation environments. The data-sets I had access to were statistically very significant, as they record live calls on the network. The challenges I faced were in characterizing the data-set's properties, amalgamating data from various sources in order to present a coherent picture of the world and enabling the first tying this in with a visualization test-bed or prototype. This dissertation uses a combination of measurement and simulation with a prototype, to evaluate various proposals for resource scheduling in particular and a modular approach that enables an end-to-end management of emergencies, in general.

Goal: Utilizing the concept of location from cell phones and addressing performance bottlenecks:

In my framework, I present call-traffic characteristics and load-balancing patterns using PCMD in two traces collected on June 7th, 2008 at the Wellington switch, which co-incided with the date on which the All Blacks played an important Rugby match in Wellington New Zealand. The data relates to information about calls placed on the country's leading telephone networks provider and is collected over the course of 24 hours with rush-hour being identified as the hour when the maximum number of calls are placed on the network. I present data pertaining to the general aspects of the number of calls placed on various cell sites during the course of the day. I further present data pertaining to the general aspects of average caller distance from the cell sites, during the course of the day. I then go on to make two hypotheses about the nature of call traffic and load balancing thereof: Load balancing can be observed on two candidate sets of cell sites (a) Busy cell-sites that have too many calls and end up shedding calls in the active set such that the originating site changes, during the call (b) Cell-sites with poor coverage that are also shedding calls.

In the case of the busiest cell sites, (that are identified to be located in downtown Wellington), observed at rush-hour, wherein the load balancing effects are studied and presented in terms of number of calls made on the busiest site and its immediate neighbours, the radius from the sites wherein the call traffic is observed in order to understand which of the neighbours are absorbing the call traffic and the proximity of the sites shedding and absorbing call traffic to the Westpac Stadium, which is where the Rugby game took place. I observe that load balancing is directional, in that the cell-site that is closer to the stadium absorbed more calls from the busiest site, rather than the cell site that was the nearest neighbour to the busiest site. In the case of the sites with poor coverage, I observe that the coverage levels of the neighbouring sites complement each other as the trace is studied alongside the time-stamp on the call that was placed on the sites in question. This suggests that when a cell-site has reduced coverage, in order to handle its calls, in the same temporal proximity, its immediate neighbours absorb calls as indicated by complementary Ec/Io levels.

From the perspective of Emergency management, understanding load balancing is important because crowds tend to form around emergency sites. Whether people are converging or diverging from a site of interest, as in the case of a game or any event with significant numbers of people involved, estimating how call traffic is handled by the underlying network enables us to propose more robust heuristics on which portions of the crowd to source for information. For example, given the choice, as an Emergency Responder, when I contact civilians on the ground, should I contact the portion on the periphery of the cell or those that are closer to the center of the cell. Arguably, the trajectory of an emergency, such as in the case of an earthquake, can be de-centralized and in this case, the use of a tower's properties, such as its inherent noise levels (as I demonstrated with a rural site) becomes relevant to estimate. Cell Breathing effects are most pertinent at certain distances from the center of the cell. This strengthens my case for estimating and utilizing a user's location correctly before choosing to contact them for more reliable information, as response times are critical in these situations and choosing a good candidate set makes a difference.

Goal: Establishing benchmarks for commercial applications

Setting aside privacy concerns for the sake of argument, when proposing augmenting a cell phone user with a reputation vector or value, one of whose tuples would be the social network of the caller I evaluated a number of social networking ideas from related research. The proposals seemed fine but a lot of them were tested *arbitrarily*, which motivated me to propose TESS.

In establishing benchmarks for Social Networks, I have presented the importance of the notion of reputation, assigning the same for cell phone users and the diversity in data-sets used in several sub fields within Social Network research, such as extracting metrics, understanding community formation, visualization and understanding trust and privacy within these structures. Understanding a user's reputation alongside location is important in order to have one more variable that speaks to the validity and utility of the information provided by the users co-located with the emergency. Data-sets are interesting and important because they ultimately lead to the creation of useful benchmarks [115], which can be used to evaluate new proposals in research. Data sets further have to represent the problem space accurately in order to validate the utility of the solutions. Social Network simulation also requires a robust understanding of these data sets. In this thesis, I have presented my confidence metric, TESS, whose various ratings elucidate whether the data used has the desired attributes. I see some variability in the average ratings, across various sub-sections of research, specifically social network extraction, community formation and trust and privacy studies.

Goal: Utilizing lessons learned to enable better Emergency Response.

I extend emergency response simulation beyond its usual purposes of training and visual analytics to introduce specialized scheduling agents, that can be used to evaluate different administrative policies and domains that act in tandem to resolve emergencies. While I borrow ideas from systems research, the number of unknowns and the time-frame over which my runs need to be completed are orders of magnitude different. Additionally, the visualization framework has to incorporate realistic renditions of the schedulers that are chosen and this poses additional challenges.

My prototype provides a user the opportunity to evaluate various agents that are computer-mediated augmented with visual simulation. To demonstrate the use of my prototype, I generated multiple scenarios and ran simulations of each with different mixes of agent characteristics implemented in their response behavior. I further modified the ways in which fine-grained resource-sharing policies can be implemented in order to provide insight into what might be useful practises for planning the procedures and policies each first responder should follow during emergency management, while catering to disparate administrative domains.

By tying in the framework presented in Chapter 6, with an actual prototype demonstrated in Chapter 7, which is written in a modular manner in order to enable the features of the framework, I demonstrate that an end-to-end management of emergencies can be enabled using different sources of information including *co-located* users.

8.2. Future Directions

While addressing performance bottle-necks, future work would include characterizing the movement of users in the area of the noisy cell sites and further understand the impact of geographical location, interference and terrain on call volume and load balancing in the underlying network. In establishing benchmarks, future work would include placing side-by-side the data-sets used in visualization tools built to be utilized in Social Network analysis, to see if there is accurate representation of these characteristics. Since the visualization section of this research is pretty advanced, with several existing tools such as Vizter, JUNG, it would validate whether the metrics presented here are adhered to well. I contend that the actual definition of the metrics such as trust, privacy, degree of centrality etc. will affect how these ratings are assigned to the data-sets. For example, if the definition of trust is Context-Specific [105] wherein a user is required to trust another user in a specific situation, a single snapshot of data (i.e, one which need not be sampled over a period of time as arguably the situation expires, past that point in time) might still be assigned a high rating, in the T(emporal) aspect.

With respect to my framework and associated prototype, Manikarnika, several paths lend themselves for discussion in future work including:

- Addressing problems related to integrating the aspects of *reputation* of a caller, rather than just their location.
- Addressing realistic models of resource-requirements changing, with respect to an emergency, in an end-to-end evaluation of the same.
- Understanding how the results would vary if the number of resources required is more than the four basic types I have taken into consideration and estimating computation complexity and run-times in that scenario.
- Extending these ideas to the field of logistics, specifically with respect to moving containers from shipments that arrive in ports, where handles are often using short-range communication devices with associated RFIDs.

8.3 Privacy Issues

One of the biggest issues in evaluating my ideas around reputation and other aspects of location information from live calls was that of privacy. Subscribers have the right to privacy and this concern has also been one of the biggest limiters in the commercial success of Location-Based Services.

1. www.gsa.com

2. http://en.wikipedia.org/wiki/Mobile_phone

3. UMTS Red Book

4. http://news.cnet.com/8301-27080_3-10398073-245.html?tag=facebook

5. Application of the stratification concept to radio networks and their gateways, Floreani and Dajej, Computer Networks and ISDN Systems, Vol 28, Issue 5, March 1996, Pages 675-687

6. Towards a Formal Standard for Interoperability in M&S/System of Systems Integration, Bernard Zeigler, Saurabh Mittal, *Proc. IEEE Int. Conf. on Networks, SICON'93* (September 1993), pp. 6–11.

7. D.J. Floreani and A.J. Dadej, An architecture for a distributed narrow-band packet radio network. *Austral. Telecomm. Res. J. (ATR)* **27** 2 (1993).

8. Flexent/Autoplex Wireless Networks, Executive Cellular Processor (ECP) PCMD and related features, Release 27, 28, Lucent Technologies

9. http://en.wikipedia.org/wiki/Haversine_formula

10. http://www.cellular-news.com/story/9674.php

11. http://www.strategyanalytics.com/default.aspx?mod=ReportAbstractViewer&a0=4736

12. http://www.pluggd.in/india-mobile-market-report-sms-as-a-vas-service-297/

13. http://www.india-cellular.com/Handset-Prices.htm

14. www.hspa.gsmworld.com

15. **http://vitalanalytics.files.wordpress.com/2009/05/game-providers-in-india.jpg**

16. http://www.eurekalert.org/pub_releases/2009-04/lp-osm042709.php

17. http://benchmarks.gomez.com

18. http://timesofindia.indiatimes.com/biz/india-business/Tata-DoCoMo-launches-per-second-billing-for-roaming-calls/articleshow/5257948.cms

19. BREW

20. http://www.fcc.gov/pshs/services/911-services/enhanced911/Welcome.html

21. http://www.nytimes.com/2005/10/02/national/nationalspecial/02ice.html?pagewanted=print

22. http://www.mobilemarketingwatch.com/category/mobile-marketing/

23. http://sciencestage.com/d/1509794/formal-models-and-game-design.html

24. "Games classification system: Teaching strategic understanding and tactical awareness Tim Hopper and Rick Bell, Caphera 66: 14-19

25. Lindley, C. A. (3 October 2003). "Game Taxonomies: A High Level Framework for Game Analysis and Design". *Gamasutra*. http://www.gamasutra.com/features/20031003/lindley_01.shtml.

26. Maja Pivec, Anni Koubek, and Claudio Dondi (2004). *Guidelines for Game-Based Learning*. Pabst Science Publishers. pp. 54–57. ISBN 1593260725.

27. Deborah P. Vossen (21 April 2004) (PDF). *The Nature and Classification of Games*. **10**. pp. 75–77. http://people.stfx.ca/dvossen/3-CV/AVANTEPaper2004.pdf.

28. http://www.reverse911.com/index.php

29. ***Visualization Research with Large Displays***, Bin Wei, Claudio Silva et. Al., 2000 IEEE Computer Graphics and Applications

30. **Mobile Landscapes: Graz in Real Time,** Carlo Ratti, Andres Sevtsuk, Sonya Huang**,** *SENSEable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA***,** Rudolf Pailer**,** *Mobilkom Austria AG & Co KG, Vienna, Austria*

31. ***CarTel: A Distributed Mobile Sensor Computing System***. Bret Hull, Vladimir Bychkovsky, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Yang Zhang, Hari Balakrishnan, and Samuel Madden, in Proc. ACM SenSys, 2006.
32. ***Visualizing Large-Scale Telecommunication Networks and Services***, Eleftherios E. Koustsofios et. al, Information Visualization Research, AT&T Labroratiries, Florham Park, N.J
33. ***A measurement study of vehicular internet access using in situ wi-fi networks*** Vladimir Bychkovsky et. al., MobiCom '06, September 24-29, 2006, Los Angeles, California, USA.
34. ***ALVIN: A system for visualizing large networks*** Davood Raiei, Stephen Curial, WWW 2005, May 10-14, 2005, Chiba Japan.
35. ***Vizster: Visualizing Online Social Networks*** Jeffrey Heer, Danah Boyd, Computer Sciences Division, UC Berkeley.
36. Visualization of real-time survivability metrics for mobile networks T.A.Dahlberg, K.R.Subramanian, MSWIM 2000, Boston MA.
37. Kuo-Chung Chu, Frank Yeong-Sung Lin"Adaptive Load Balancing of Cellular CDMA Systems Considering Non-uniform Traffic Distributions ", Lecture Notes in Computer Science, 2005.
38. H. Zhu, T. Buot, R. Nagaike, and S. Harmen, ―Load Balancing in WCDMA systems by adjusting pilot power, Wireless Personal Multimedia Communications. The 5th International Symposium, Vol. 3, pp. 936 – 940, 27-30 Oct.
39. 2002. S. K. Das, S. K. Sen, and R. Jayaram, ―Handoff Triggering and Network Selection Algorithms for Load-Balancing Handoff in CDMA-WLAN Integrated Networks‖, Proceedings of EURASIP Journal on Wireless Communications and Networking Volume 2008 (2008), Article ID 136939, 14 pages.
40. http://www.allblacks.co.nz/index.cfm?layout=2008results
41. VALENT PROSPECT
42. **Long-term and session-specific user preferences in a mobile recommender system,** Quang Nhat Nguyen , Francesco Ricci , International Conference on Intelligent User Interfaces archive

    Proceedings of the 13th international conference on Intelligent user interface, 2008

43. **A privacy preserving web recommender system, Baraglia et. Al,** Proceedings of the 2006 ACM symposium on Applied computing
44. **Tag-aware recommender systems by fusion of collaborative filtering algorithms,** Fortaleza, Ceara, Brazil, Proceedings of the 2008 ACM symposium on Applied computing.
45. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, The EigenTrust Algorithm for Reputation Management in P2P Networks, In Proceedings of the Twelfth International World Wide Web Conference, 2003.,
46. Papaioannou, G. Thanasis, Stamoulis, D. George. Effective Use of Reputation in Peer-to-Peer Environments
47. Castelfranchi, Cristiano and Falcone, Rino. Social Trust: A Cognitive Approach,
48. Abdul-Rahman, Alfarez, Hailes, Stephen. Supporting Trust in Virtual Communities. *Hawaii International Conference on System Sciences 33.* Jan 2000
49. Ariely, Dan and Simonson, Itamar. Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions. *Forthcoming in Journal of Consumer Psychology.*
50. Ba, Sulin and Pavlou, Paul A.. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly.* Sep 2002
51. The Sybil Attack by John R. Douceur,
52. Sybil Attacks Against Mobile Users: Friends and Foes to the Rescue by Daniele Quercia and Stephen Hailes. IEEE INFOCOM 2010.,
53. On the establishment of distinct identities in overlay networks by Rida A. Bazzi and Goran Konjevod,
54. The Sybil Attack in Sensor Networks: Analysis & Defenses
55. http://www.csd.uoc.gr/~hy558/papers/sybilguard.pdf
56. Marchionini, G., Nolet, V., Williams, H., Ding, W., Josephus Beale, J., Rose, A., Gordon, A., Enomoto, E., Harbinson, L.: Content + connectivity = community :\ digitalresourcesforalearningcommunity. In:DL097: ProceedingsofthesecondACMinternationalconferenceonDigitallibraries,NewYork,NY,USA ,ACM(1997)212−−220
57. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69(2) (Feb 2004) 026113

58. Heer, J., Boyd, D.: Vizster: Visualizing online social networks. In: INFOVIS ’05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, Washington, DC, USA, IEEE Computer Society (2005) 5
59. Hamasaki, M.: Community focussed social network extraction. In: CSCW ’04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2004) 328–331
60. Ortiz, J.A., Tapia, A., Rangel, E.M.: Recruiting diverse, high-skilled it employees through existing virtual social networks. In: SIGMIS CPR ’06: Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research, New York, NY, USA, ACM (2006) 4–11
61. Counts, S., Geraci, J., Geraci, J.: Incorporating physical co-presence at events into digital social networking. In: CHI ’05: CHI ’05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2005) 1308–1311
62. Chung, K.S.K., Hossain, L., Davis, J.: Individual performance in knowledge intensive work through social networks. In: SIGMIS-CPR ’07: Proceedings of the 2007 ACM SIGMIS CPR conference on 2007 computer personnel doctoral consortium and research conference, New York, NY, USA, ACM (2007) 159–167
63. Matsuo, Y., Mori, J., Hamasaki, M., Ishida, K., Nishimura, T., Takeda, H., Hasida, K., Ishizuka, M.: Polyphonet: an advanced social network extraction system from the web. In: WWW ’06: Proceedings of the 15th international conference on World Wide Web, New York, NY, USA, ACM (2006) 397–406
64. danah michele boyd: Friendster and publicly articulated social networking. In: CHI ’04: CHI ’04 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2004) 1279–1282
65. Yu, B., Singh, M.P.: Searching social networks. In: AAMAS ’03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM (2003) 65–72
66. Culotta, A., Bekkerman, R., McCallum, A.: Extracting social networks and contact information from email and the web. In: Proceedings of CEAS-04, the 1st Conference on Email and Anti-Spam. (2004)

67. Gardner, T.: Visual query interfaces for wiki's and blogs. In: SIGGRAPH '04: ACM SIGGRAPH 2004 Web graphics, New York, NY, USA, ACM (2004) 5
68. Eagle, N., Pentland, A.S.: Reality mining: sensing complex social systems. Personal Ubiquitous Comput. 10(4) (2006) 255–268
69. Zachary, W.W.: A survey of augmented reality. Journal of Anthropological Research 33 (1977) 452–473
70. Knuth, D.E.: The Stanford GraphBase: A Platform for Combinatorial Computing. Addison-Wesley (1993)
71. Newman, M.E.J.: Mixture models and exploratory analysis in networks. Phys. Reviews 74 (2006)
72. Girvan, M., Newman, M.E.J. Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826
73. D. Lusseau, K. Schneider, O.J.B.P.H.E.S., Dawson, S.M.: Behavioral Ecology and Sociobiology 54. Addison-Wesley (2003)
74. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 us election. WWW-2005 Workshop on the Weblogging Ecosystem (2005)
75. Krebs, V.: Books on us politics sold by amazon.com. V. Kerbs, personal website, www.orgnet.com (2008)
76. Newman, M.E.J.: Community centrality. Phys. Reviews 74 (2006)
77. et. al, A.M.: Measurement and analysis of online social networks. Internet Measurement Conference (2007)
78. Breiger, R.L., M, K., Pattison, P.: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. National Academies Press (2003)
79. Freeman, L.: Finding Social Groups: A Meta-Analysis of the Southern Women Data. (2003) 39–45
80. John Resig, Santosh Dawara, C.M.H., Teredesai, A.: Extracting social networks from instant messaging populations. In: Workshop on Link Analysis and Group Detection (LinkKDD2004). (2004)
81. Adamic, L.A.: Friends and neighbours on the web. http://www.sciencedirect.com/science/article/B6VD1-48Y6SYK-1/2/8216f71528567bedb3a194828fe68c34

82. Goecks, J., Mynatt, E.D.: Leveraging social networks for information sharing. In: CSCW ’04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2004) 328–331
83. Khan, J.I., Shaikh, S.: Relationship algebra for computing in social networks and social network based applications. In: WI ’06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA, IEEE Computer Society (2006) 113–116
84. Zhou, D., Ji, X., Zha, H., Giles, C.L.: Topic evolution and social interactions: how authors effect research. In: CIKM ’06: Proceedings of the 15th ACM international conference on Information and knowledge management, New York, NY, USA, ACM (2006) 248–257
85. McCallum, A.: Information extraction, data mining and joint inference. In: KDD ’06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2006) 835–835
86. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: SIGIR ’07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2007) 103–110
87. Andrade, N., Mowbray, M., Lima, A., Wagner, G., Ripeanu, M.: Influences on cooperation in bittorrent communities. In: P2PECON ’05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, New York, NY, USA, ACM (2005) 111–115
88. Cai, D., Shao, Z., He, X., Yan, X., Han, J.: Mining hidden community in heterogeneous social networks. In: LinkKDD ’05: Proceedings of the 3rd international workshop on Link discovery, New York, NY, USA, ACM (2005) 58–65
89. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In: KDD ’06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2006) 44–54
90. M¨uller-Prove, M.: Community experience at openoffice.org. interactions 14(6) (2007) 47–48

91. Pfeil, U.: Online social support for older people. SIGACCESS Access. Comput. (88) (2007) 3–8
92. Millen, D.R., Patterson, J.F.: Stimulating social engagement in a community network. In: CSCW ’02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2002) 306–313
93. Gr, K.: Ubiquitous hypermedia and social interaction in physical environments. In: HYPERTEXT ’06: Proceedings of the seventeenth conference on Hypertext and hypermedia, New York, NY, USA, ACM (2006) 119–120
94. Qamra, A., Tseng, B., Chang, E.Y.: Mining blog stories using community-based and temporal clustering. In: CIKM ’06: Proceedings of the 15th ACM international conference on Information and knowledge management, New York, NY, USA, ACM (2006) 58–67
95. Tantipathananandh, C., Berger-Wolf, T., Kempe, D.: A framework for community identification in dynamic social networks. In: KDD ’07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2007) 717–726
96. Freeman, L.: Visualizing social networks. Journal of Social Structure
97. de Nooy, W., Mrvar, A., Batagelj, V.: Exploratory Social Network Analysis with Pajek. Cambridge University Press, New York, NY, USA (2004)
98. ´Alvaro Reis Figueira, Laranjeiro, J.B.: Interaction visualization in web-based learning using igraphs. In: HT ’07: Proceedings of the 18th conference on Hypertext and hypermedia, New York, NY, USA, ACM (2007) 45–46
99. Schulz, H.J., Nocke, T., Schumann, H.: A framework for visual data mining of structures. In: ACSC ’06: Proceedings of the 29th Australasian Computer Science Conference, Darlinghurst, Australia, Australia, Australian Computer Society, Inc. (2006) 157–166
100. Berry, M., Hamilton, M.: Mobile computing, visual diaries, learning and communication: changes to the communicative ecology of design students through mobile computing. In: ACE ’06: Proceedings of the 8th Austalian conference on Computing education, Darlinghurst, Australia, Australia, Australian Computer Society, Inc. (2006) 35–44
101. Chung, W., Leung, A.: Supporting web searching of business intelligence with information visualization. In: WI ’07: Proceedings of the IEEE/WIC/ACM International

Conference onWeb Intelligence,Washington, DC, USA, IEEE Computer Society (2007) 807–811

102. Domingos, P., Richardson, M.: Mining the network value of customers. In: KDD ’01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2001) 57–66

103. Chen, H., Atabakhsh, H., Tseng, C., Marshall, B., Kaza, S., Eggers, S., Gowda, H., Shah, A., Petersen, T., Violette, C.: Visualization in law enforcement. In: dg.o2005: Proceedings of the 2005 national conference on Digital government research, Digital Government Research Center (2005) 229–230

104. Saltz, J.S., Hiltz, S.R., Turoff, M.: Student social graphs: visualizing a student’s online social network. In: CSCW ’04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2004) 596–599

105. Quasthoff, M., Sack, H., Meinel, C.: Why https is not enough – a signature-based architecture for trusted content on the social web. In: WI ’07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA, IEEE Computer Society (2007) 820–824

106. O’Donovan, J., Smyth, B.: Trust in recommender systems. In: IUI ’05: Proceedings of the 10th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2005) 167–174

107. Kaufman, J.H., Edlund, S., Ford, D.A., Powers, C.: The social contract core. In:WWW ’02: Proceedings of the 11th international conference on World Wide Web, New York, NY, USA, ACM (2002) 210–220

108. Gefen, D.: Reflections on the dimensions of trust and trustworthiness among online consumers. SIGMIS Database 33(3) (2002) 38–53

109. Sabater, J., Sierra, C.: Social regret, a reputation model based on social relations. SIGecom Exch. 3(1) (2002) 44–56

110. Fullam, K.K., Barber, K.S.: Learning trust strategies in reputation exchange networks. In: AAMAS ’06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM (2006) 1241–1248

111. Cvrcek, D., Kumpost, M., Matyas, V., Danezis, G.: A study on the value of location privacy. In: WPES ’06: Proceedings of the 5th ACM workshop on Privacy in electronic society, New York, NY, USA, ACM (2006) 109–118

112. Chatfield, C., Hexel, R.: Privacy and community connectedness: designing intelligent environments for my cities. In: OZCHI ’07: Proceedings of the 2007 conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artifacts and environments, New York, NY, USA, ACM (2007) 265–272

113. Yu, H., Kaminsky, M., Gibbons, P.B., Flaxman, A.: Sybilguard: defending against sybil attacks via social networks. In: SIGCOMM ’06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications, New York, NY, USA, ACM (2006) 267–278

114. Gibson, R.: Who’s really in your top 8: network security in the age of social networking. In: SIGUCCS ’07: Proceedings of the 35th annual ACM SIGUCCS conference on User services, New York, NY, USA, ACM (2007) 131–134

115. Narayanan, R., ¨ Ozisikyilmaz, B., Zambreno, J., Memik, G., Choudhary, A.N.:Minebench: A benchmark suite for data mining workloads. In: IISWC. (2006) 182–188

116. Carl A. Waldspurger. **Lottery and Stride Scheduling: Flexible Proportional-Share Resource Management**, Ph.D. dissertation, Massachusetts Institute of Technology, September 1995. Also appears as Technical Report MIT/LCS/TR-667.

117. Greedy Packet Scheduling on Shortest Paths (1991), by Yishay Mansour , Boaz Patt-Shamir Journal of Algorithms

118. A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem (2006), by Rubén Ruiz , Thomas Stützle European Journal of Operational Research

119. A Greedy Algorithm for a Time-Dependent Scheduling Problem (2002), by Stanislaw Gawiejnowicz Wieslaw , Wieslaw Kurc , Lidia Pankowska , Adam Mickiewicz

120. Ito, A., Murakami, H., Watanabe, Y., Fujii, M., Yabe, T., Haraguchi, Y., Tomoyasu, Y., Kakuda, Y., Ohta, T., Hiramatsu, Y.: ***Universal use of information delivery and display system using ad hoc network for deaf people in times of disaster***. (2008) 486–491

121. Wu, W., Cao, J., Zheng, Y., Zheng, Y.P.: ***Waiter: A wearable personal healthcare and emergency aid system.*** (2008) 680–685

122. Lee, R.G., Chen, K.C., Hsiao, C.C., Tseng, C.L.: ***A mobile care system with alert mechanism.*** Information Technology in Biomedicine, IEEE Transactions on 11(5) (2007) 507–517

123. Inoue, Y., Sashima, A., Ikeda, T., Kurumatani, K.: ***Indoor emergency evacuation service on autonomous navigation system using mobile phone.*** (2008) 79–85

124. Filippoupolitis, A., Hey, L., Loukas, G., Gelenbe, E., Timotheou, S.: ***Emergency response simulation using wireless sensor networks.*** In: Ambi-Sys '08: Proceedings of the 1st international conference on Ambient media and systems, ICST, Brussels, Belgium, Belgium, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2008) 1–7

125. Chitumalla, P., Harris, D., Thuraisingham, B., Khan, L.: ***Emergency response applications: Dynamic plume modeling and real-time routing***. Internet Computing, IEEE 12(1) (2008) 38–44

126. Basu, S.N.: ***Fermi 2 emergency response information system.*** Energy conversion, IEEE transactions on EC-1(2) (1986) 16–22

127. Serpa, D.P., Walker, D.M., Jenckes, T.A.: ***Emergency monitoring, assessment and response system for diablo canyon nuclear power plant.*** Nuclear Science, IEEE Transactions on 28(1) (1981) 236–241

128. Burnett, B., Bowering, G., Hagelstrom, R., Tate, D., Ramanujan, R.: ***Multi-radio nodes for mobile mesh networks for emergency response and battle command systems.*** (2007) 1–7

129. Symons, L., Pavia, R., Hodges, M.: ***Emergency response in national marine sanctuaries.*** (2005) 345–351 Vol. 1

130. Gertman, D.I., Bruemmer, D.J., Scott Hartley, R.: ***Improving emergency response and human- robotic performance.*** (2007) 334–340

131. Bansal, A., Patel, K., Gupta, G., Raghavachari, B., Harris, E., Staves, J.: ***Towards intelligent services: a case study in chemical emergency response.*** (2005) –758

132. Kung, H.Y., Ku, H.H., Wu, C.I., Lin, C.Y.: ***Intelligent and situation-aware pervasive system to support debris-flow disaster prediction and alerting in taiwan.*** J.Netw. Comput. Appl. 31(1) (2008) 1–18

133. Miyauchi, K., Yonezawa, Y., Ogawa, H., Maki, H., Caldwell, W.: ***A mobile phone-based safety and life support system for elderly people.*** (2005) 81–84

134. Pries, R., Hobfeld, T., Tran-Gia, P.: ***On the suitability of the short message service for emergency warning systems.*** Volume 2. (2006) 991–995

135. Drane, C., Macnaughtan, M., Scott, C.: P***ositioning gsm telephones.*** Communications Magazine, IEEE 36(4) (1998) 46–54, 59

136. Scott Shane, E.L.: ***Stumbling storm-aid effort put tons of ice on trips to nowhere.***The New York Times (2005)

APPENDIX I – Mobile Phone Prices and Features.

NOKIA - N86	Rs.24,000	Internal Memory- 8 GB ;Memory Card Type-microSD,TransFlash ;Extensible Memory-16 GB	240 x 320 pixels	Available
NOKIA - N85	Rs.22,000	Internal Memory- 74 MB ;Memory Card Type-MicroSD ;Extensible Memory- 8 GB	240 x 320	WAP 2.0 XHTML, HTML
SAMSUNG - M8800 Pixon	Rs.22,000	Internal Memory- 200 MB ;Memory Card Type-Micro SD ;Extensible Memory- 8 GB	240 x 400	WAP 2.0, xHTML, HTML
SONYERICSSON - W980i	Rs.22,000	Internal Memory- 8 GB	240 x 320 pixels	WAP 2.0/xHTML, HTML (NetFront), RSS reader
SAMSUNG - i780	Rs.17,600	Internal Memory- 128 MB RAM, 256 MB ROM ;Memory Card Type-microSD (TransFlash)	320 x 320 pixels	WAP 2.0 / HTML, Opera Mobile 8.65
MOTOROLA - Q9 h	Rs.17,500	Internal Memory- 96 MB RAM, 256 MB ROM ;Memory Card Type-Micro SD ;Extensible Memory- 8 GB	320 x 240 pixels	HTML
NOKIA - E71	Rs.17,000	Internal Memory- 110 MB storage, 128 MB RAM ;Memory Card Type-Micro SD ;Extensible Memory- 8 B	320 x 240 pixels	WAP 2.0/xHTML, HTML
SAMSUNG - Beat DJ M7603	Rs.15,500	Internal Memory-50 MB ;Memory Card Type-microSD, TransFlash ;Extensible Memory- 2 GB	240 x 400 pixel	Available
NOKIA - N79	Rs.15,000	Internal Memory-50 MB ;Memory Card Type-microSD;Extensible Memory- 4 GB	240 x 320	WAP 2.0/xHTML
SONYERICSSON - W705	Rs.15,000	Internal Memory- 120 MB ;Memory Card Type-Memory Stick Micro (M2),;Extensible Memory- 8 GB	240 x 320 pixels	WAP 2.0/HTML, RSS
SAMSUNG - Star 3G	Rs.12,000	Internal Memory- 80 MB ;Memory Card Type-	WQVGA	WAP 2.0, HTML

		microSD;Extensible Memory- 8 GB		
SONYERICSSON - W595i	Rs.12,000	Internal Memory- 40 MB ;Memory Card Type-Memory Stick Micro (M2);Extensible Memory- 8 GB	240 x 320 pixels	WAP 2.0/xHTML, HTML (NetFront)
NOKIA - 6208 Classic	Rs.10,500	Internal Memory- 22 MB ;Memory Card Type-microSD;Extensible Memory- 8 GB	320 x 240	WAP 2.0/xHTML
SONYERICSSON - C510	Rs.10,200	Internal Memory- 100 MB ;Memory Card Type-Memory Stick Micro (M2);Extensible Memory- 8 GB	240 x 320 pixels	WAP 2.0/HTML (NetFront), RSS
NOKIA - 5320 XpressMusic	Rs.10,100	Internal Memory- 140 MB ;Memory Card Type-microSD;Extensible Memory- 8 GB	240 x 320	WAP 2.0/xHTML, HTML
SONYERICSSON - W580i	Rs.9,200	Internal Memory- 12 MB ;Memory Card Type-Memory Stick Micro ;Extensible Memory- 2 GB	240x320 pixel	WAP 2.0/HTML(NetFront), RSS feeds
MOTOROKR - U9	Rs.9,000	Internal Memory- 25 MB ;Memory Card Type-microSD	240x320	
SONYERICSSON - W302	Rs.7,350	Internal Memory- 20 MB ;Memory Card Type-Memory Stick Micro ;Extensible Memory- 4 GB	176 x 220 pixels	WAP 2.0/xHTML
NOKIA - 6303 Classic	Rs.7,200	Internal Memory- 17 MB ;Memory Card Type-microSD,TransFlash ;Extensible Memory- 4 GB	240 x 320 pixels	Available
NOKIA - 3120 Classic	Rs.7,190	Internal Memory-24 MB ;Memory Card Type-microSD,TransFlash ;Extensible Memory- 8 GB	240 x 320 pixels	WAP 2.0/xHTML, HTML
SAMSUNG - J800	Rs.7,100	Internal Memory- 18 MB ;Memory Card Type-microSD,TransFlash	176 x 220 pixels	WAP 2.0/xHTML, HTML
NOKIA - 3610 Fold	Rs.6,850	Internal Memory-30 MB ;Memory Card Type-microSD;Extensible	240 x 320 pixels	WAP 2.0/xHTML, HTML

		Memory-4 GB		
NOKIA - 6300	Rs.6,800	Internal Memory-7.8 MB ;Memory Card Type-microSD;Extensible Memory-2 GB	240 x 320 pixels	
SAMSUNG - L700	Rs.6,800	Internal Memory-30 MB ;Memory Card Type-microSD	128 x 160	WAP 2.0/xHTML, HTML
SAMSUNG - C 5212	Rs.6,800	Internal Memory-60 MB ;Memory Card Type-microSD;Extensible Memory-8 GB	176 x 220 pixels	WAP 2.0/xHTML
SONYERICSSON - F305	Rs.6,700	Internal Memory-10 MB ;Memory Card Type-Memory Stick Micro (M2) ;Extensible Memory-4 GB	176 x 220 pixels	Available
SONYERICSSON - S302	Rs.6,600	Internal Memory-15 MB ;Memory Card Type-MemoryStick Micro	176 x 220 pixels	Available
MOTOYUVA - A810	Rs.6,200	Memory Card Type-MicroSD;Extensible Memory-2 GB	240 x 320	WAP 2.0/xHTML, HTML
SAMSUNG - METRO 3500	Rs.5,399	Internal Memory-30 MB ;Memory Card Type-MicroSD ;Extensible Memory-8 GB	240 x 320 pixels	
NOKIA - 7210 Supernova	Rs.5,300	Internal Memory-30 MB ;Memory Card Type-MicroSD ;Extensible Memory-2 GB	240 x 320	
SAMSUNG - METRO 3310	Rs.4,849	Internal Memory-30 MB	2.1 Inches QVGA resolution	WAP 2.0 , GPRS , EDGE
NOKIA - 6085	Rs.4,700	Internal Memory- 4 MB ;Memory Card Type-MicroSD ;Extensible Memory-2 GB	128 x 160 pixels	WAP 2.0/xHTML
SAMSUNG - B 2100 Marine	Rs.4,649	Internal Memory- 7 MB ;Memory Card Type-MicroSD ;Extensible Memory-8 GB	1280x1024 pixels	

SONYERICSSON - W200i	Rs.4,400	Internal Memory-27 MB ;Memory Card Type-Memory Stick Micro ;Extensible Memory-2 GB	640x480 pixels	WAP 2.0/xHTML, HTML(NetFront), RSS feeds
NOKIA - 2680 Slider	Rs.4,200	Internal Memory-32 MB	128 x 160	